

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 15, 2004, 19:19:09 ; Search time 33 Seconds
(without alignments)
3971.533 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQWTRATEPRYP.....WDEPKPLLCQSYTLSDSE 2517

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13215	100.0	2517	1 NCR2_HUMAN	Q9Y618 h nuclear r
2	10987.5	83.1	2472	1 NCR2_MOUSE	Q9WU42 mus musculus
3	4187	31.7	2453	1 NCR1_MOUSE	Q60974 mus musculus
4	4147.5	31.4	2440	1 NCR1_HUMAN	Q75376 homo sapien
5	714.5	5.4	533	1 NCR1_RAT	Q9WUB5 rattus norv
6	576	4.4	3644	1 MINT_MOUSE	Q62504 mus musculus
7	569.5	4.3	3664	1 MINT_HUMAN	Q98T58 homo sapien
8	526	4.0	2142	1 BAT2_MOUSE	P48634 homo sapien
9	503	3.8	2716	1 OSA_DROME	Q8IN94 drosophila
10	492.5	3.7	2774	1 MAPA_RAT	P34926 rattus norv
11	470.5	3.6	5262	1 MLL2_HUMAN	O14686 homo sapien
12	466.5	3.5	5560	1 SPEN_DROME	O88X83 drosophila
13	462.5	3.5	5085	1 PCLO_RAT	Q9JK86 rattus norv
14	451.5	3.4	2805	1 MAPA_HUMAN	P78559 homo sapien
15	439.5	3.3	2067	1 NCO6_MOUSE	Q9J119 m nuclear r
16	439	3.3	4911	1 MLL3_HUMAN	Q8HEZ4 homo sapien
17	438.5	3.3	1780	1 YKZ6_MOUSE	P34333 caenorhabdi
18	434.5	3.3	5038	1 PCLO_MOUSE	Q9QYX7 mus musculus
19	432	3.3	2167	1 SHK1_RAT	Q9WV48 rattus norv
20	428.5	3.2	2715	1 MLL4_HUMAN	Q9UNM6 homo sapien
21	428	3.2	2063	1 NCO6_HUMAN	O14686 h nuclear r
22	424.5	3.2	2464	1 MAPB_MOUSE	P14873 mus musculus
23	415.5	3.1	1226	1 SNT1_YEAST	P25357 saccharomyc
24	410.5	3.1	5147	1 PCLO_HUMAN	Q9Y6V0 homo sapien
25	410	3.1	2468	1 MAPB_HUMAN	P46821 homo sapien
26	407	3.1	5120	1 PCLO_CHICK	Q9PU36 gallus gall
27	406.5	3.1	2459	1 MAPB_RAT	P15205 rattus norv
28	400.5	3.0	1464	1 CA11_HUMAN	P02452 homo sapien
29	399	3.0	2161	1 SHK1_HUMAN	Q9Y566 homo sapien
30	399	3.0	4903	1 MLL3_MOUSE	Q8BRH4 mus musculus
31	396	3.0	3703	1 ABP1_HUMAN	Q15911 homo sapien
32	391.5	3.0	1411	1 TCOF_HUMAN	Q13428 homo sapien
33	391.5	3.0	1460	1 CA11_CANPA	Q9XSJ7 canis famil

RESULT 1

ALIGNMENTS

34	391	3.0	3924	1	ANK2_HUMAN	Q01484 homo sapien
35	390	3.0	2944	1	CA17_HUMAN	Q02388 homo sapien
36	389.5	2.9	3866	1	HRX_MOUSE	P55200 mus musculus
37	389	2.9	1670	1	CA34_HUMAN	Q01955 homo sapien
38	382.5	2.9	1362	1	BRD4_HUMAN	O60885 homo sapien
39	380	2.9	3969	1	HRX_HUMAN	Q03164 homo sapien
40	379	2.9	1685	1	CA54_HUMAN	P29400 homo sapien
41	378	2.9	1183	1	DRPL_RAT	P54258 rattus norv
42	375.5	2.8	3726	1	ABP1_MOUSE	Q61329 mus musculus
43	373	2.8	1763	1	CA24_ASCSU	P27393 ascaris suu
44	368	2.8	1185	1	DRPL_HUMAN	P54259 homo sapien
45	364	2.8	1426	1	BCL9_HUMAN	O00512 homo sapien
46	360	2.7	1509	1	BCL9_HUMAN	Q9H2M4 homo sapien
47	358	2.7	1618	1	NEST_HUMAN	P48681 homo sapien
48	355.5	2.7	2842	1	APC_RAT	P70478 rattus norv
49	355	2.7	2090	1	N214_HUMAN	P35658 homo sapien
50	353	2.7	1453	1	CA11_CHICK	P02457 gallus gall
51	353	2.7	1983	1	TF20_MOUSE	Q9EPQ8 mus musculus
52	352	2.7	1838	1	CA15_HUMAN	P20908 homo sapien
53	350.5	2.7	1736	1	CA2B_HUMAN	P13942 homo sapien
54	350.5	2.7	2843	1	APC_HUMAN	P25054 homo sapien
55	349.5	2.6	1822	1	ZAP3_HUMAN	P49750 homo sapien
56	348.5	2.6	1464	1	CA13_MOUSE	P08121 mus musculus
57	348	2.6	1669	1	CA14_MOUSE	P02463 mus musculus
58	348	2.6	3256	1	KI67_HUMAN	P46013 homo sapien
59	346	2.6	1338	1	ACIN_MOUSE	Q9JIX8 mus musculus
60	344	2.6	2845	1	APC_MOUSE	O61315 mus musculus
61	341.5	2.6	1026	1	NFH_HUMAN	P12036 homo sapien
62	341	2.6	1453	1	CA11_MOUSE	P11087 mus musculus
63	341	2.6	1815	1	SHK3_RAT	Q9JLU4 rattus norv
64	340	2.6	1466	1	CA13_HUMAN	P02461 homo sapien
65	339.5	2.6	1341	1	ACIN_HUMAN	Q9UKV3 homo sapien
66	339.5	2.6	1690	1	CA44_HUMAN	P53420 homo sapien
67	339	2.6	1960	1	TF20_HUMAN	Q9UGU0 homo sapien
68	338.5	2.6	1505	1	CUT2_HUMAN	O14529 homo sapien
69	337.5	2.6	2426	1	SON_HUMAN	P18583 homo sapien
70	335.5	2.5	5703	1	MUSB_HUMAN	Q9HC84 homo sapien
71	335	2.5	1516	1	CA1H_HUMAN	P30060 homo sapien
72	334	2.5	1739	1	DOTL_HUMAN	Q8TEK3 homo sapien
73	332.5	2.5	1902	1	SMF1_HUMAN	O14497 homo sapien
74	332.5	2.5	2404	1	SON_MOUSE	Q9GX47 mus musculus
75	332	2.5	810	1	NFM_BOVIN	O77788 bos taurus
76	330.5	2.5	1418	1	CA12_HUMAN	P02458 homo sapien
77	330.5	2.5	2349	1	TPR_HUMAN	P12270 homo sapien
78	330	2.5	1367	1	AMVH_YEAST	P08640 saccharomyc
79	330	2.5	1461	1	IE18_PRTIF	P11675 pseudorabie
80	329.5	2.5	915	1	NFM_HUMAN	P07197 homo sapien
81	329.5	2.5	1083	1	T2D3_HUMAN	O00268 homo sapien
82	325	2.5	1003	1	MBD6_HUMAN	Q96DN6 homo sapien
83	325	2.5	1781	1	AK12_HUMAN	Q02952 homo sapien
84	324	2.5	3149	1	TEGU_EBV	P03186 epstein-bar
85	323.5	2.4	1603	1	CA1F_HUMAN	Q07092 homo sapien
86	323	2.4	1446	1	IE18_PRTVA	P33479 pseudorabie
87	322.5	2.4	2004	1	MYS3_HUMAN	Q92794 homo sapien
88	320.5	2.4	1669	1	CA14_HUMAN	P02462 homo sapien
89	319.5	2.4	1362	1	CA21_CHICK	P02467 gallus gall
90	319	2.4	1736	1	CA2B_MOUSE	Q64739 mus musculus
91	317	2.4	555	1	GP1_CHLRE	Q9FPG6 chlamydomon
92	316	2.4	1603	1	PSC_DROME	P35820 drosophila
93	315.5	2.4	1459	1	CA12_MOUSE	P28481 mus musculus
94	315.5	2.4	1496	1	CA25_HUMAN	P05997 homo sapien
95	315	2.4	6632	1	UN89_CABEL	O01761 caenorhabdi
96	314	2.4	1255	1	MUC1_HUMAN	P15941 h mucin 1 p
97	312.5	2.4	1712	1	CA24_HUMAN	P08572 homo sapien
98	311	2.4	2688	1	ZEP1_MOUSE	Q03172 mus musculus
99	310.5	2.3	3358	1	PGCV_MOUSE	Q62059 mus musculus
100	309	2.3	1087	1	NFH_MOUSE	P19246 mus musculus

QY	1410	AHEGLVATVKEAGRSIHEI	PREELRHTELP	PLAPRPLKEGSI	TGQTP	LYDTGASTGSK	1469
Db	1380	THEGVVATVKEAGRSIHEI	PREELRRTELP	PLAPRPLKEGSI	TGQTP	LYDSCAPSTGK	1439
QY	1470	KHDVRSIGSPGRTPFPV	PLDVNMADARALERAC	YEESLSKSRPGT	ASSSGS	TARGAPVI	1529
Db	1440	KHDVRSIGSPGRFPFAL	PDIMADARALERAC	YEESLSKSRGT	SSGAGGS	ITRGAPVV	1499
QY	1530	VPELGKPROSLTYEDHC	APFAGHLPRGSPVTME	PTPRLOEGSLSSK	ASQDRKLTSTP		1589
Db	1500	VPELGKPROSLTYEDHC	APTSHLPGSPVTME	PTPRLOEGSLSSK	ASQDRKLTSTP		1559
QY	1590	REIAKSPHSTVPEHHPH	PISPYEHLRLRGVGD	LYRSHIPIAFDPT	SIPRGIPLD	AAAA	1648
Db	1560	REIAKSPHSTVPEHHPH	PISPYEHLRLRGVGD	LYRSHIPIAFDPT	SIPRGIPLEAAAA		1619
QY	1649	YYLPHLAPNTYPHLYP	PYIRGYDPTAALENQ	TINDYITSCQMHNT	TATAMAQAD		1708
Db	1620	YYLPHLAPNTYPHLYP	PYIRGYDPTAALENQ	TINDYITSCQMHNA	ASNAQAD		1679
QY	1709	MURGISPRESSLALN	YAAGPGIIDL	SOVPHLPVLVPT	PGTATAMDRLAYL	PTAPOP	1768
Db	1680	MURGISPRESSLALN	YAAGPGIIDL	SOVPHLPVLVPT	PGTATADRLAYL	PTADPP	1739
QY	1769	SSRHSSPLSPGPHLTK	PTTSSSERDRDRDR	EREKSKILTSIT	TTHVHAPLWR		1828
Db	1740	SSRHSSPLSPGPHLTK	PTTSSSERERERER	ERD-----	KSILTSITTHVHAPLWR		1793
QY	1829	PGTEQSSGSGSGGSG	SSRSPASHAHQHS	PISPRTOALQORP	SVLHNTGKGIIT		1888
Db	1794	PGTEQSSGA-----	GSSSPASHT--	HQSPISPRTOALQ	ORPSVLHNTSMKGVT		1843
QY	1889	AVEPSKPTVL---	RTSTSSPVRAATFP	PATHCPGLGTG	LVGYVTLMPEPVL	LKPEAPRV	1945
Db	1844	SVEPGTIVLWARST	STSSPVRAATFP	PATHCPGLGTG	LVGYVTLMPEPVL	LKPKETSRV	1903
QY	1946	ARPERPATGHAFLAK	PARGLEPASSPSK	SEPRPLVPVSGH	ATARTPAKNLAPH		2005
Db	1904	ARPERARVDAGHAFL	TKPPGR---	EPASSPSKSEPR	SLAPSSHTAIRT	PAKNLAPH	1960
QY	2006	HASPDPAASADPHRE	KTSKPSIOELRLS	LGYH-GSSYPGE	VEPVSPVSPSL		2064
Db	1961	HASPDPA--PTASD	LHREKTSKPSIO	ELRLSILGYHSG	AGYSPDGVETIP	SPVSPSL	2019
QY	2065	THDKGLPHELDKSHL	SELRKQPGPVKLG	GEAAHLPHLRPL	PEPSSPILLQAP		2124
Db	2020	THDKGLSKPLEELEK	SHLELRKQPGPMK	LSEAAHLPHLRPL	PEPSSPILLQAP		2079
QY	2125	GVKGHQVVTLAQHI	SEVITQDTRHH	POOLGAPLAPLY	SPFGASCPVLDL	RRPPSDLY	2184
Db	2080	GKGHQVVTLAQHI	SEVITQDTRHH	POOLGAPLAPLY	SPFGASCPVLDL	RRPPSDLY	2139
QY	2185	LPPPDHGAPARGSPH	SEGGKRSPEPKNT	SVLGGEDGIEP	VPDPEGTEPFGH	SRNAVPL	2244
Db	2140	LPPPDHGTARGSPH	SEGGKRSPEPKNT	SVLGGEDAIEP	VPDPEGTEPFGH	SRNAVPL	2199
QY	2245	LYRDEGOTESRMGSK	SPGNTSQPPAF	SKLTESNAMYK	SKQENKLNTHNR	NEPY	2304
Db	2200	LYRDEGQGEF--RM	GLESFGNTSQP	PTTFKLTESN	AMYKSKQENKLN	THNRNEPY	2258
QY	2305	NISOPGTEIFNMPA	ITGTLMTVRSQAV	QEHASTNMGLE	AIIRKALMGKYD	OWEESPPLS	2364
Db	2259	NIGQFGTEIFNMPA	ITAGLMTCSQAV	QEHASTNMGLE	AIIRKALMGKYD	OWEPPPLG	2318
QY	2365	ANAFNPLNASASL	P--AAMPTIADGR	SDHTYLSFGGGG	KKAKVSCRSSR	KKAKSPAGLAS	2423
Db	2319	ANAFNPLNASASL	PAAAMPITADGR	SDHALTFPGGGG	KKAKVSCRSSR	KKAKSPAGLAS	2378
QY	2424	GDRPPSVSVHSEGD	QNRRTPLNRVWED	RPSAGSTFP	PNPLIMELQAGV	MASSPPPG	2483
Db	2379	GDRPPSVSVHSEGD	QNRRTPLNRVWED	RPSAGSTFP	PNPLIMELQAGV	MASSPPPG	2438

2484 LPAGSPLAGPHAWDEBP

2439 LAAGSPLAGPHAWDEBP

RESULT 3

NCRI MOUSE

ID NCRI MOUSE STANDARD; PRT: 2453 AA.

AC Q60974; Q60812; Rel. 40, Created

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 42, Last annotation update)

DE Nuclear receptor co-repressor 1 (N-COR1) (Retinoid X receptor interacting protein 13) (RIP13).

DE NCOR1 OR RXRIP13.

GN Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).

RC TISSUE=Pituitary;

RX MEDLINE=96008539; PubMed=7566114;

RA Hoerlein A.J., Naeae A.M., Heinzel T., Torchia J., Gloss B., Kurokawa R., Ryan A., Kamei Y., Soederstroem M., Glass C.K., Rosenfeld M.G.;

RT "Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor.";

RL Nature 377:397-404 (1995).

RN [2]

RP SEQUENCE OF 1792-2453 FROM N.A. (ISOFORM LONG).

RC TISSUE=Liver;

RX MEDLINE=95280959; PubMed=7760852;

RA Seol W., Choi H.S., Moore D.D.;

RT "Isolation of proteins that interact specifically with the retinoid X receptor: two novel orphan receptors.";

RL Mol. Endocrinol. 9:72-85 (1995).

RN [3]

RP INTERACTION WITH HDAC7.

RX MEDLINE=20442375; PubMed=10984530;

RA Downes M., Ordentlich P., Kao H.-Y., Alvarez J.G.A., Evans R.M.;

RT "Identification of a nuclear domain with deacetylase activity.";

RL Proc. Natl. Acad. Sci. U.S.A. 97:10330-10335 (2000).

CC -!- FUNCTION: Mediates the transcriptional repression activity of some nuclear receptors by promoting chromatin condensation, thus preventing access of the basal transcription.

CC -!- SUBUNIT: Interacts with HDAC7. Forms a large corepressor complex that contains SIN3A/B and histone deacetylases HDAC1 and HDAC2. This complex associates with the thyroid (TR) and the retinoid acid receptors (RAR) in the absence of ligand.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- ALTERNATIVE PRODUCTS: Event=Alternative splicing; Named isoforms=2;

CC Name=Long;

CC IsoId=Q60974-1; Sequence=Displayed;

CC Name=Short;

CC IsoId=Q60974-2; Sequence=VSP_003411;

CC -!- TISSUE SPECIFICITY: Ubiquitous.

CC -!- DOMAIN: The N-terminal region contains repression functions that are divided into three independent repression domains (RD1, RD2 and RD3). The C-terminal region contains the nuclear receptor-interacting domains that are divided in two separate interaction domains (ID1 and ID2).

CC -!- DOMAIN: The two interaction domains (ID) contain a conserved sequence referred to as the CORNR box. This motif is required and sufficient for binding to unliganded TR and RARs. Sequences flanking the CORNR box determine nuclear hormone receptor specificity.

CC -!- SIMILARITY: Contains 1 SANT-A domain.

CC -!- SIMILARITY: Contains 1 Myb-like domain.

CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS FAMILY.

1477 IGSRTFFPPVHPLDWMAD-ARALERACYE-----ESLKRPGTASSSGSIARGAPVIV 1530
 1418 ITGPKL--PRGMEIIPENIKVVERKEDVAGEFVARHVSVVSSGFSVLST--L 1472
 1531 PELCKPQSPPLYEDHGA-----PFAGHLPRGSPVTWREPTPLQEGSLSSK-ASQDRK 1584
 1473 HEAPKAQLSPGLYDSSARRTPVSQNTISRGSPMNR-----TSDVSSSSKASHERK 1525
 1595 LTSTPRE-----IAKSPHSTVEHPHPHPISPYEHLLRGVGVLDLYRSHIPLAAPTSPRG 1640
 1526 STLTPTQRESIPAKSPGVGVNDPIVSH--SPDPHRSAAAGEVVRSHLPHLDP-AMPFH 1582
 1641 IPLDAAAYLPHLAPNTPVPHLYPYLIRGYDPTAALEN-ROTIINDYITSQOMHNT 1699
 1593 RALDDPAAYLQRLSPFTGPGVPSQYLY-----AMENTRQTIINDYITSQOMQVNL 1633
 1700 ATAMAQRAADMLRGLSPRESSALALNYAAGPRGIIDLSQVPHLPVLVPTPGTATMDLA 1759
 1634 -----REDVTGLSPREQPLGLPYA-TRGIIDLTNMP-TELPHAGGTSTPFMDRIT 1685
 1760 YLTPAQPFSSR-HSSSPLSGGPHLTKPTTSSRERDRDRDRERERKSILTST 1818
 1686 YTPGTQVTFPPPPYNAALSPGPHL---AAASAERERERERERERERERER 1742
 1819 TTVEHAP---IWRPGTESGSSGSGSGSGSSRPSHSHAHQHSPIPTQDA-LQOR 1874
 1743 ERIAAPADLYLRPGSEQ-----PGRPGSHGVVRSPSP-SVRTETILQOR 1787
 1875 PSVLNHTGMKGIITAVEPSKPTVLSTSTSPV-----RPAATFPATHCPLGGTLDGVVP 1930
 1788 PSVFOCTNGTSVITLDPQAQRLMPLPSGGGSIQGLPAGRYNTAADA-LAALVDAAS 1846
 1931 TLMPEVLLKEAPR-----VARPERPRADTGHAFKAPPARSGLEPASSPSK 1977
 1847 APQMDVSKTESKHEARLEENLRSSAAVSQQLQKNLEVEKRSVQCVCTSSALPSG 1906
 1978 GSEPRPLV-----PPVSGHATARTAPKN--LAPHASDPDPAPPASADPHRE 2024
 1907 KAQPHASVVVSEAGDKGPPKSRVEELRTGKTITTAANFIDVITRQIASDKARER 1966
 2025 KTSKFSIQEILRSLGVHSGSYSPGVEPVSSPSLTHDKGLPKHLEBLDKSHLEG 2084
 1967 GSQSSDS-----SSLSHRYETASDAIEVISPASSPAPQEKQAYQVDMVKANQAE 2020
 2085 ELRKPQPGPKLGEAAHLPLR-----PLPSQSSSPLLQ--APGVKGHQRVTLAQ 2137
 2021 ESTROYEGP-----LHHYRSQCESPSQPPQPLPPSSQSGMGQVPRTHRLITLAD 2071
 2138 HISEVITQDYTRH-POOLSAPL-----PAPLYSFGACSPVLDLRRPPSDLYLPPPD--- 2189
 2072 HICQITQDFARNQVPSQASTSTFTQSPSALSSTP-----VRKTSRYSPPSQSOT 2123
 2190 --HGAPA-RGSPH-----SEGGKRSPEPNKTSVLGGEDGIEPVSPPEGMTPEG-HSRSA 2240
 2124 VLHRRPGVRSPENLVDKSRGSRGKSPERSHI---PSEYEPISPPQ--PAVHEKQD 2177
 2241 VYPLYRGDQTEPSRMGKSPGNTOPPAFFKLTESNAMSVMYKSKQKINKLNTHNEN 2300
 2178 SMLLSQKQVDPABQRSDSRSPGISYLPFFTKL-ESTSPMWKSKQKIFRKLNSGGG 2236
 2301 EPEYNISQGTIEFNMPAITGTGLMVTYSQAOVESHASTNNGLEAIRKALMGKYDQWEE- 2359
 2237 DSDMAAQPGTEIENLPAVITSGVSRSHSFADPAS-NLGLIEDIIRKALMGSDFKVED 2295
 2360 -----SPPLSANAFNPLNASASLPAAMPITTAAGRSHTITSPCGG-GKAKVSGRPSRK 2413
 2296 HGVVMSHEV---GIMPGSASTV-----VTSSEARDEGEFSPHAGVCKPKLINKNSRK 2347
 2414 AKSPAPGLA--SGDRPPSVSVHSGDCNRTPLTNVWEDRPSAGSTFPFYNPLIMRL 2471
 2348 SKSPFPGQSYLGTERPSSVSVHSGDYHROTP--GNAWEDRPSSTGSTQFPYNPLTIRM 2405

2472 QAGVMASPPPLGDPAGSPL--AGPH---HAWDEEPKPLLCQSYETLSDSE 2517
 2406 ----LSSTPTQIACAPSAITQAPHQNRWEREPALLSAQYETLSDSD 2452

RESULT 4
 NCRL_HUMAN
 ID NCRL_HUMAN STANDARD; PRT; 2440 AA.
 AC 075376; Q9UPV5; Q9UQ18;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Nuclear receptor co-repressor 1 (N-COR1) (N-COR).
 GN NCOR1 OR KIAA1047.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=98393736; PubMed=9724795;
 RA Wang J., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.,
 RT "ETO, fusion partner in t(8;21) acute myeloid leukemia, represses
 RT transcription by interaction with the human N-COR/MSIN3/HDAC1
 RT complex.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).
 RN [2]
 RP SEQUENCE OF 782-2440 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:197-205(1999).
 RN [3]
 RP SEQUENCE OF 974-2440 FROM N.A.
 RX MEDLINE=99375328; PubMed=10444336;
 RA Nagaya T., Chen K.-S., Fujieda M., Ohmori S., Richer J.K.,
 RA Horwitz K.B., Lupski J.R., Seo H.;
 RT "Localization of the human nuclear receptor co-repressor (hN-COR) gene
 RT between the CMT1A and the SMS critical regions of chromosome
 RT 17p11.2.";
 RL Genomics 59:339-341(1999).
 CC -!- FUNCTION: Mediates the transcriptional repression activity of some
 CC nuclear receptors by promoting chromatin condensation, thus
 CC preventing access of the basal transcription.
 CC -!- SUBUNIT: Interacts with HDAC7 (By similarity). Forms a large
 CC corepressor complex that contains SIN3A/B and histone deacetylases
 CC HDAC1 and HDAC2. This complex associates with the thyroid (TR) and
 CC the retinoid acid receptors (RAR) in the absence of ligand.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- DOMAIN: The N-terminal region contains repression functions that
 CC are divided into three independent repression domains (RD1, RD2
 CC and RD3). The C-terminal region contains the nuclear receptor-
 CC interacting domains that are divided in two separate interaction
 CC domains (ID1 and ID2).
 CC -!- DOMAIN: The two interaction domains (ID) contain a conserved
 CC sequence referred to as the CORNR box. This motif is required and
 CC sufficient to permit binding to unliganded TR and RARs. Sequences
 CC flanking the CORNR box determine nuclear hormone receptor
 CC specificity.
 CC -!- SIMILARITY: Contains 1 SANT-A domain.
 CC -!- SIMILARITY: Contains 1 MYB-like domain.
 CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	EMBL; AF044209; AAC33550.1; -	
DR	EMBL; AB028970; BA882999.1; -	
DR	EMBL; AB019524; BAA75814.1; -	
DR	TRANSPAC; T04687; -	
DR	Genew; HGNC:7672; NCOR1.	
DR	MIM; 600849; -	
DR	GO; GO:0003714; P:transcription co-repressor activity; TAS.	
DR	GO; GO:0006366; P:transcription from Pol II promoter; TAS.	
DR	InterPro; IPR001005; Myb_DNA_binding.	
DR	Pfam; PF00249; myb_DNA-binding; 2.	
DR	SMART; SM00717; SANT; 2.	
DR	PROSITE; PS50090; MYB 3; 1.	
KW	Nuclear protein; Transcription regulation; DNA-binding; Repressor; Coiled coil.	
FT	DOMAIN 174 216	COILED COIL (POTENTIAL).
FT	DOMAIN 254 312	INTERACTION WITH SIN3A/B.
FT	DOMAIN 299 328	COILED COIL (POTENTIAL).
FT	DOMAIN 437 482	SANT-A (POTENTIAL).
FT	DNA_BIND 620 670	MYB.
FT	DOMAIN 501 557	COILED COIL (POTENTIAL).
FT	DOMAIN 607 617	PRO-RICH.
FT	DOMAIN 988 1816	INTERACTION WITH ETO.
FT	DOMAIN 2055 2059	CORN BOX OF ID1.
FT	DOMAIN 2263 2267	CORN BOX OF ID2.
FT	DOMAIN 58 64	POLY-GLN.
FT	DOMAIN 593 603	POLY-ALA.
FT	DOMAIN 1032 1035	POLY-PRO.
FT	DOMAIN 1707 1712	POLY-ALA.
FT	DOMAIN 1952 1963	POLY-SER.
FT	CONFLICT 1014 1014	L-> V (IN REF. 2).
FT	CONFLICT 1508 1509	PP -> SS (IN REF. 2).
FT	CONFLICT 1561 1561	W -> R (IN REF. 2).
FT	CONFLICT 1567 1567	Q -> H (IN REF. 2).
SQ	SEQUENCE 2440 AA; 270263 MW; 60AA4D7964D00EDAB CRC64;	

430	Qy	DB	VYQDRQFMVWTDHEKEIFKDKFIQHPNFGLIASYLERSKVPDCVLVYYLTTKKNENYKA	489
482	Qy	Qy	LVRRSY--RRRGKSQQQQQQQQQQQQQQQMPRSPSOEKKDEKEKEKEBEKEBEKEVEN	540
490	Db	Db	LVRENYCKRRNQOIAIPSQBEKVEKEE--DKAEKTEKBEKEKDEBEKDEKUSKEN	547
541	Qy	Qy	DKEDLLKEKTDGTSGRONDEKEAVASKGRKTANSQGRKGRITRSMANEANSSEAITPOQ	600
548	Db	Db	TKE---KDKIDGTA-EETEBREQATPRGRTANSQGRKGRITRSMTEANASAAAAA	603
601	Qy	Qy	S-----AELASMLNSESRTWTEEMETAKGLLEHGRNWSIARMPVSKTYSOCK	650
604	Db	Db	TEPPPPPLPPPPPEISTPVSRTWTEEMEVAKGLVEHGRNWAIAKMWGTKSQAQK	663
651	Qy	Qy	NFYFNKKRQNLDEILQOHKLKWEKERNARKKKKAPAAASEAAFPVVVEDEMEASGV	710
664	Db	Db	NFYFNKRNHLNLLQOHKQKTSRPREERDVQSCEVASVTA-----QEDEDIAAS--	717
711	Qy	Qy	SGNEBEMVBEAALHASGNEVPRGCSGPATVNNSDTESIPSPHTEAAKDTQONGPKPP	770
718	Db	Db	--NEEENPEDSE-----VEAVK-----	732
771	Qy	Qy	ATLGADGPPGPPTPPRTSRAPIEPTPASEATGATPPAP--PSFS-APPVVVPKEEKE	828
733	Db	Db	-----PSDSPENATSRGNTEPAVELE---PTTETAPSTSPSLAVPSTKPAEDES	779
829	Qy	Qy	EETAAAAPV-----EEGEQKPPAAEELAVDTGKAEBEPVKSECTEEAE	871
780	Db	Db	VETQVNDISIABTAEBQMDVDOQSHSABEGVCDPPATK--ADSDVDVVRVPPENHASKVE	837
872	Qy	Qy	EGPAKGDAABAEATAGALKAEKGGGSRATTAKSSGA-----PQ-DSDSATCSADEV	926
838	Db	Db	GDNTKERDLORA-----SEKVPRDEDLVVAQIQAORPEQPSDNDSSATCSADE--	887
927	Qy	Qy	DEABEGDKNRLSP---RPSLLTPTGDPRANASPOK--PLDLKQLKQRAAIIPII-----	976
888	Db	Db	-DVDGEPERQMFPMDSKPSLLNPTGSGILV--SSPLKPNPLDLPLQOLHRAAVIPMVSCPT	945
977	Qy	Qy	-----QVTKVHEPPREDAAATPKPAPPAPPQNIQPSDAPQPGSSP	1019
946	Db	Db	CNPIGTVPVSGYALYQBHIKAMHESALLE-----EQRQEQEIDLCECRSSTSP	993
1020	Qy	Qy	RGKSRSPAPPADKFAAEAAQKLPDGPCCWTSGLPFPPVPPREVIKASHPADPSFASYAP	1079
994	Db	Db	CGTSKSP-----NRE-----W-----EVLQAPAH-----QLITNLP	1019
1080	Qy	Qy	PGHPLPLGLHDHTARVLP,PRPPTTSNPPPLISSAKHPSVLERQIGALISQ-----MSVQLH	1134
1020	Db	Db	EGVRLP-----TTRTPRPPPLIPSSKTTVASEK-ESFTI--MGSGISQGTGGTYLTSHQ	1071
1135	Qy	Qy	VPYSEHAKAP--VGPVTMGLPLPMDPKKLAPFSGVKQEOQLSPROAGPPRESLGVPTAQEAS	1193
1072	Db	Db	ASTQETPKESVGSISLGLPRQOESAKSATLPYIKQEEFSPKSNQSPQEGLLV--RAQHEG	1130
1194	Qy	Qy	VLRTGALGVPGGSIKTIPISTVPSPDSAITYRGSIHTGTPA-----DVLTKGTITR	1245
1131	Db	Db	VWRGTA-GAIQEGSITRGTPTSKISVESIPSLRGSITQGTPALPQTGIPTREALVKGSIR	1189
1246	Qy	Qy	IIGEDSPRLDGRDRLSPKHVIVYKKGHVLSYEGGMSVTQCSKEDGRSSSGPPHETA	1305
1190	Db	Db	MPIEDSSP--EKGREAAAGKHVIVYKKGSHILSYDNIKNA-----REGTRSPRTAHEIS	1242
1306	Qy	Qy	APKRTYDMEGRVGRAS-----SASISGLMGRATPPRRHSPH-HLKEQHHRGSIITOG	1358
1243	Db	Db	L-KRSVESVGNTKQGMWRESVPVAPLGLICRALP--RGSPHSLKERTVLGSGIMOG	1299
1359	Qy	Qy	IPRSYVBEAQDYLRRBEAKLLKREGTPPPPPPPSRDLTEAYKTOALGPLKLPKPAHEGLVATV	1418
1300	Db	Db	TPRATTESFDGL-KYPKQIKRES-----PPTRAFEGAI-----TKGCP-YDG-ITTI	1344
1419	Qy	Qy	KERAGRSIHETPREEL-----RHTPELPLAPRLKEGSIITQGTPLKYDGTGASTGSKHDV	1473
1345	Db	Db	KEMGRSITHEIPRODILTQSRKTPVQVQSTRTPIEGSISQGTPIKFDNN--SGOSATKNV	1403

QY	1474	RSLIGSGRTGPPVHPLDVMAD-ARALERACYE-----ESLKRPGCTASSGSGSIARGAP	1527	QY	2462	FPYNPLMRLOAGVMSPPPLPAGSG-PLAGPH---HAWDEPKPLLCQYETLSDE	2517
DB	1404	KSLITGSKLGRGPPLEIIPENIKVVERKYEDVAGETVRSRHTSVVSSGSLVLRST-	1462	DB	2383	FPYNPLTWRM---LSSTPTPIACAPSAVNOAAPHQNRIVEREPAPILLSAQYETLSDS	2439
QY	1528	VIVPELAKGPROSLTYEDHGA-----PPAGHLPRGSPVMTWREPTPLRQESLSSKASQD	1582				
DB	1463	--LHEAPKAQLSPGIVYDTSARRTPVSYQNTMSRGSPMMNRTSDVTIP-----FNKSTNHE	1516				
QY	1593	RKLSTPRE-----TAKSPHSVPEHHHPHPISPYBHLHLRGVGVDLVRSHPILAFDPTSP	1638				
DB	1517	RKSTLTPTQRESIIPAKSPVPGVDVSH--SPFDPHRRGAGTAGEVYVSHLPTQLDP-AMP	1573				
QY	1639	RGIPLD-AAAAAYILPHLAPNPTPHLYPPYLRGYPDTAALEN-ROTIINDYITSQOMH	1696				
DB	1574	FHRALDPAAYLQFQLSPPTGYPQYQLY-----AMENTROTILNDYITSQOMQ	1624				
QY	1697	HNTATAMQADMLRGLSPRESSLALNYAAGPRGIIIDLSQVPHLPVLVPPTPGTPATAMD	1756				
DB	1625	VNL-----RPDVARGLSPREQLPLPYPA-TRGLIDLTNMPP-TILVPHPGTSTPPMD	1676				
QY	1757	RLAYLPTAPQFSSR-HSSPLSPGCPHLLAKPTTTSSSERDRDRDRDREREKSIIL	1815				
DB	1677	KITLIPGTCITFPFPYNSASMSFGPHLL-----AAAASAEEREREREKERERER---IA	1730				
QY	1816	TSTTTVEHAPIWRPCTEQSSGSGSGGGSSSRPASHAHQHSPISTQDA-LQOR	1874				
DB	1731	RASSDL-----YLKPGSEQ-----PCRPGRSHGYVRGSP-SVETQETMLQOR	1771				
QY	1875	PSVLHNTGMKGIITAVESEKPTVLRSTSTSSPV-----RPAATFPFATHCPGLGTLGYP	1930				
DB	1772	PSVFGCTNGTSTVITLDPDTAQLRIMPLPAGGSPISQGLPASYNTAADA-LAALVDAAS	1830				
QY	1931	TLMEPVLLPK-----EAPRVARPERPAD-----TGHAFIAX	1962				
DB	1831	APQMDVSKTESKHEAARLEENLRSAVSEQQOQLEKTEVEKRSVQCLYTSAPFSG	1890				
QY	1963	PPARSGLEPASS-----PSKSEBRPLVPVSGHATARTPAKN--LAPHHASDPDPAPAS	2017				
DB	1891	KP-----QPHSSVVVSEAGDKGP--PPKSRVEELRTGKTTITAFIDVILTRQIAS	1943				
QY	2018	ASDPHREKTSQKPFSTQELSLRSLGSHGSSYSGPEGVPSVSPSLTHDKGLPKHLDEL	2077				
DB	1944	DKDAREGQSDDSS-----SSLSSHRVETPDSALEVISPASSPAPQOKLTQYQPEV	1997				
QY	2078	DKSHLEGLRKPQGPVKLGGEANHLPHLPSPESQPSLLQAPG-----VKGH	2129				
DB	1998	KANQAENDPTROVEGP-----LHHYRP--QOESPSPQOQLPPSQAEGMGQVPT	2045				
QY	2130	ORVVTLAQHISEVITDITRHH-----POOLSAPLPAPLYSPFGA--SCPVLDLRRPPSD	2182				
DB	2046	HRUITLADHLICQITQDFARNQVSSQTPOQ--PPTSTFQNSPALSALVSTPV--RTKTSN	2099				
QY	2183	LYLPPFD-----HGAP-ARGSPH-----SEGGKRSPEPKNTSVLGGEGDIEPVSPPEGM	2231				
DB	2100	RYPESQAQSVHQRPGSRVSPENLVDRSGRSPGKSPERSHV---SSFPYFPISPPQ--	2154				
QY	2232	TEPHGSRSAVYPLLYRDEQTEPSRMGSKSPGNTSQPPAFPSKLTESNAMVSKKQKQIN	2291				
DB	2155	VPVVEHQDLSLLLSQSGAPAEQRNDARSPGISVILPSPFTKL-ENTSPWKSKQEIF	2213				
QY	2292	KKLNTNRNPEYVNIQPGTEIFNMPAITGTGLMTRVRSQAVQBSHATNMGLEAIIIRKALM	2351				
DB	2214	RKLNSGGGSDMAAAGPGTEIFNLPAVTTSGSVSRGHSFADPAS-NLGLEDIIRKALM	2272				
QY	2352	GKYDQWEE-----SPPLSANAFNPLNASLASLPAAMPITAADGRSHDLTSPGGG--KA	2403				
DB	2273	GSFDRKVEDHGVNMSQPMGV-----VPGTANTSV-----VTSGETRREEGDPSHSGGVCKP	2324				
QY	2404	KVSGRPSRKAAPSAP--GLASGRPPSPSVSHSEGCNRRTPPLTNRWEDRPSSAGSTP	2461				
DB	2325	KLISKNSRKSIPGQYLGTERPSPSVSHSEGDYHRQTP--GWAWEDRPSSTGSGTQ	2382				

RESULT 5

ID	NCR1	RAT	STANDARD;	PRT;	533	AA.
AC	Q9WUE5;	O70463;				
DT	16-OCT-2001	(Rel. 40, Created)				
DT	16-OCT-2001	(Rel. 40, Last sequence update)				
DT	10-OCT-2003	(Rel. 42, Last annotation update)				
DE	Nuclear receptor co-repressor 1 (N-COR1)	(Fragment).				
GN	NCOR1.					
OS	Rattus norvegicus (Rat).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
OX	NCBI_TaxID=10116;					
RN	[1]	SEQUENCE FROM N.A.				
RP	TISSUE=Brain;					
RC	MEDLINE=99371771; PubMed=10441327;					
RX	Boutell J.M., Thomas P., Neal J.W., Weston V.J., Duce J., Harper P.S.,					
RA	Jones A.L.;					
RT	"Aberrant interactions of transcriptional repressor proteins with the					
RL	Huntington's disease gene product, huntingtin.";					
RN	Hum. Mol. Genet. 8:1647-1655 (1999).					
RP	[2]	SEQUENCE OF 476-528 FROM N.A.				
RC	TISSUE=Skeletal muscle;					
RX	MEDLINE=99421707; PubMed=10491148;					
RA	Schuler M.J., Buehler S., Pette D.;					
RT	"Effects of contractile activity and hypothyroidism on nuclear hormone					
RL	receptor mRNA isoforms in rat skeletal muscle.";					
CC	Eur. J. Biochem. 264:982-988 (1999).					
CC	FUNCTION: Mediates the transcriptional repression activity of some					
CC	nuclear receptors by promoting chromatin condensation, thus					
CC	preventing access of the basal transcription (By similarity).					
CC	SUBUNIT: Interacts with HDAC7. Forms a large corepressor complex					
CC	that contains SIN3A/B and histone deacetylases HDAC1 and HDAC2.					
CC	This complex associates with the thyroid (TR) and the retinoid					
CC	acid receptors (RAR) in the absence of ligand (By similarity).					
CC	SUBCELLULAR LOCATION: Nuclear (By similarity).					
CC	DOMAIN: The N-terminal region contains repression functions that					
CC	are divided into three independent repression domains (RD1, RD2					
CC	and RD3). The C-terminal region contains the nuclear receptor-					
CC	interacting domains that are divided in two separate interaction					
CC	domains (ID1 and ID2).					
CC	DOMAIN: The two interaction domains (ID) contain a conserved					
CC	sequence referred to as the CORNR box. This motif is required and					
CC	sufficient to permit binding to unliganded TR and RARs. Sequences					
CC	flanking the CORNR box determine nuclear hormone receptor					
CC	specificity.					
CC	SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS					
CC	FAMILY.					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	EMBL; AF124821; AAC32566.1; --					
DR	EMBL; AF059311; AAC14567.1; --					
DR	InterPro; IPR001005; Myb DNA binding.					
DR	PROSITE; PS50090; MYB 3; PARTIAL.					
KW	Nuclear protein; Transcription regulation; DNA-binding; Repressor.					
FT	NON_TER	1				POLY-SER.
FT	DOMAIN	48				CORNR BOX OF ID1.
FT	DOMAIN	153				157
FT	DOMAIN	357				CORNR BOX OF ID2.

```
FT CONFLICT 484 484 R -> W (IN REF. 2).
SQ CONFLICT 497 497 A -> V (IN REF. 2).
ST SEQUENCE 533 AA; 57794 MW; 7DF60F8228227EC2 CRC64;

Query Match
  5.4%; Score 714.5; DB 1; Length 533;
Best Local Similarity 38.5%; Pred. No. 1.4e-16; Gaps 24;
Matches 205; Conservative 68; Mismatch 192; Indels 67;

QY 2016 ASADPHREKTKQSPFSIQEELRLSLGVHSSYSPEGVFPVSPVSLTHDKGLPKHLE 2075
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 38 ASDKADRGSSQSDSS-----SSLSHRYEAPDAIEVISPASSAPPQKPKQTYQPE 91
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 2076 ELDKSHLEGELPKQPGPKVLGGEAHLPHLR-----PLPESQPSSSLPLQTPG---VK 2127
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 92 MYKANQAEENESQQYEGP-----LTHYSQQGSPSPQOQPLPPSSQ-AEGMGQVP 141
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 2128 GHORVVTLAQLHSEVITDTRHHPQQLSAPLAPLYSPFGA--SCPVLDRRPPSDLYL 2185
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 142 RTHRLITLADHLCIQTQDFARN--QVPSQPTSTFTQSPSALSTPV---RTKPSRYS 196
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 2186 PPPD-----HGAPA-RGSPH-----SEGGKSPPEPKTSLVLCGGEDGIEPVSPPEGMTPE 2234
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 197 PESQSQTVLHPRPGRPVSRPENLVKSRGSRPGKSPERSHI---PSEVPETSPQG---P 250
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 2235 G-HRSRAVPLLYRGOETEPERMKGKPGNTSQPPAFKSLTENSAMVSKKQEIINKK 2293
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 251 AVHERQDSMLLISQRGMDPAEQRSRSPGISYLPYFTKL-ESTSPMVKSKQEIIRK 309
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 2294 LINTNRNPEYNISQPTIEFNWPAITGTGLMYRSQAOEASHTMGLAIIRKALMGK 2353
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 LNSGGGSDMAAQFGTEIFNLPAVTTSGAVSSRSHSFADPAS--NLGLEDIIRKALMG 368
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 2354 Y-DWEESPPILSANAFNPLNASASLPAAMPITAADGRSDHTLTSPPGG-GRKAVGRPSS 2411
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 369 PDKVEDHGVMPHPVGVVPGSASTSV---VTSSSTRDEGDPSPHSGVCKPLINKNS 425
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 2412 RKAQSPAG--LASGDRPPSVSSVHSEGDGNNRTPLTRVWEDRPSSAGSTPPFYNPLIM 2469
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 426 RKSKSPIQGNVLTGTRPSSVSSVHSEGDYHRTPE--GWAMEDRPSSGTSTQFPYNPLTI 483
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 2470 RLQAGVMAAPPPGGLPAGSG-PLAGPH---HAWDEEPKLLCSQYETLSDSE 2517
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 484 RM---LSSTPTPTTACAPSAITQAAAPHQOOSRIWEREPAPLISAQYETLSDSD 532
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 6

```
MINT_MOUSE
ID MINT_MOUSE STANDARD; PRT; 3644 AA.
AC Q62504; Q80TN9; Q99PS4; Q9QZW2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN MINT OR SHARP OR KIAA0929.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE OF 49-3644 FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR
RP LOCATION, TISSUE SPECIFICITY, DNA-BINDING, AND INTERACTION WITH MSX2.
RC TISSUE-Testis;
RX MEDLINE=93379811; PubMed=10451362;
RA Newberry E.P., Latifi T., Towler D.A.;
RT "The RRM domain of MINT, a novel mxx2 binding protein, recognizes and
RT regulates the rat osteocalcin promoter.";
RL Biochemistry 38:10678-10690(1999).
[2]
RN SEQUENCE OF 1-112 FROM N.A.
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,
RA Hashizume W., Hayashida K., Hirozane T., Hori F., Imotani K.,
RA Ishii Y., Itoh M., Kagawa I., Kawai J., Kojima Y., Kondo S., Konno H.,
RA Koya S., Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,
```

```
RA Ohno M., Ohsato N., Saito R., Sakazume N., Sano H., Sasaki D.,
RA Sato K., Shibata K., Shiraki T., Tagami M., Takeda Y., Waki K.,
RA Watahiki A., Muramatsu M., Hayashizaki Y.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 69-3644 FROM N.A. (ISOFORM 2), AND VARIANTS THR-348;
RP PHE-762; PHE-773 AND LEU-933.
RC STRAIN=ICR; TISSUE=Brain;
RA Sakamoto T., Gotou T., Isagawa Y., Mimura H., Kimura K., Kawaichi M.;
RT "MINT/spen negatively regulates Notch signaling by inhibiting RBP-
RT J/Su(H) activity";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE OF 318-578 FROM N.A.
RP TISSUE=Cochlea;
RX MEDLINE=97237053; PubMed=9119401;
RA Crozet P., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA Hamet C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA Weil D., Fajol R., Petit C.;
RT "Cloning of the genes encoding two murine and human cochlear
RT unconventional type I myosins.";
RL Genomics 40:332-341(1997).
[5]
RN SEQUENCE OF 2598-3644 FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
[6]
RN TISSUE SPECIFICITY.
RP MEDLINE=22261914; PubMed=12374742;
RX Oswald F., Kostecka U., Astrahantseff K., Bourteelle S., Dillinger K.,
RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoeschel W., Liptay S.,
RA Schmid R.M.;
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT pathway.";
RL EMBO J. 21:5417-5426(2002).
[7]
RN FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=22483652; PubMed=12594956;
RA Kuroda K., Han H., Tani S., Tanigaki K., Tun T., Furukawa T.,
RA Taniguchi Y., Kurooka H., Hamada Y., Toyokuni S., Honjo T.;
RT "Regulation of marginal zone B cell development by MINT, a suppressor
RL of Notch/RBP-J signaling pathway.";
IMMUNITY 18:301-312(2003).
CC -!- FUNCTION: Essential corepressor protein, which probably regulates
CC different key pathways such as the Notch pathway. Negative
CC regulator of the Notch pathway via its interaction with RBPSUH,
CC which prevents the association between NOTCH1 and RBPSUH, and
CC therefore suppresses the transactivation activity of Notch
CC signaling. Blocks the differentiation of precursor B cells into
CC marginal zone B cells. Probably represses transcription via the
CC recruitment of large complexes containing histone deacetylase
CC proteins. May bind both to DNA and RNA.
CC -!- SUBUNIT: Interacts with NCOR2, HDAC1, HDAC2, RBBP4, MED3 and
CC MTA1L1. Interacts with the nuclear receptors RAR and PPARG.
CC Interacts with RAR in absence of ligand. Bind to the steroid
CC receptor RNA coactivator SRA (By similarity). Interacts with MSX2.
CC Interacts with RBPSUH; this interaction may prevent the
CC interaction between RBPSUH and NOTCH1.
CC -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q62504-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q62504-2; Sequence=VSP_008564;
CC Note=No experimental confirmation available;
```


QY 2409 PSSRKAKSPAGLASGRPP-----SYSSVHSEGDGNCNRTPLTN-RVWEDRPPSAGSTPF 2462
 DB 3247 PSSQLQGLFL-----TPPVVTHGVQIVHSSGELFQERYVGDVTRTYHAPAOQLTHQTQF 3299

QY 2463 PY-NPLMLRLQAGWASPPPPGLPAGSGPLAGP 2494
 DB 3300 PVASSISLASRTKTSQAQVPEGEPLQSTQSAQP 3332

RESULT 7

MINT HUMAN STANDARD; PRT: 3664 AA.

AC Q96T58; Q9H9A8; Q9NH5; Q9UQ01; Q9Y556;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
 GN MINT OR SHARP OR KIAA0929.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1] SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,
 RP RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3;
 RP RAR AND MTA1L1.
 RC TISSUE=Liver, and Pituitary;
 RX MEDLINE=1231190; PubMed=11331609;
 RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,
 RA Hon M., Evans R.M.;
 RA "Sharp, an inducible cofactor that integrates nuclear receptor
 RT repression and activation.";
 RL Genes Dev. 15:1140-1151(2001).
 RN [2] SEQUENCE FROM N.A.
 RP Bird C.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 RN [3] SEQUENCE OF 294-3664 FROM N.A.
 RA Rhodes S., Huckle E.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [4] SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.
 RP TISSUE=Embryo, and Teratocarcinoma;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
 RA "NEDO human cDNA sequencing project.";
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [5] SEQUENCE OF 2002-3664 FROM N.A.
 RP TISSUE=Brain;
 RX MEDLINE=9246063; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. XIII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RN [6] J. Biol. Chem. 268:11867-11874 (1993).
 RP INTERACTION WITH PP4R.
 RX MEDLINE=21874127; PubMed=11867749;
 RA Shi Y., Hon M., Evans R.M.;
 RA "The peroxisome proliferator-activated receptor delta, an integrator
 RT of transcriptional repression and nuclear receptor signaling.";
 RN [7] Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618(2002).
 RP FUNCTION, AND INTERACTION WITH RBP4.
 RX MEDLINE=22261914; PubMed=12374742;
 RA Oswald F., Kostezka U., Astrahantseff K., Bourteele S., Dillinger K.,

RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
 RA Schmid R.M.;
 RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
 RT pathway.";
 RN [8] EMBO J. 21:5417-5426(2002).
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF SPOC DOMAIN.
 RX MEDLINE=22777836; PubMed=12897056;
 RA Ariyoshi M., Schwabe J.W.R.;
 RA "A conserved structural motif reveals the essential transcriptional
 RT repression function of Spen proteins and their role in developmental
 RT signaling.";
 RN [9] Genes Dev. 17:1909-1920(2003).
 CC -1- FUNCTION: Essential corepressor protein, which probably regulates
 CC different key pathways such as the Notch pathway. Negative
 CC regulator of the Notch pathway via its interaction with RBP4, and
 CC which prevents the association between NOTCH1 and RBP4, and
 CC therefore suppresses the transactivation activity of Notch
 CC signaling. Blocks the differentiation of precursor B cells into
 CC marginal zone B cells. Probably represses transcription via the
 CC recruitment of large complexes containing histone deacetylase
 CC proteins. May bind both to DNA and RNA.
 CC -1- SUBUNIT: Interacts with MSX2 (By similarity). Interacts with
 CC NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and MTA1L1. Interacts with
 CC RBP4; this interaction may prevent the interaction between
 CC RBP4 and NOTCH1. Interacts with the nuclear receptors RAR and
 CC PP4R. Interacts with RAR in absence of ligand. Bind to the
 CC steroid receptor RNA coactivator SRA.
 CC -1- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
 CC -1- TISSUE SPECIFICITY: Expressed at high level in brain, testis,
 CC spleen and thymus. Expressed at intermediate level in kidney,
 CC liver, mammary gland and skin.
 CC -1- INDUCTION: By hormone 17-beta-estradiol (E2).
 CC -1- DOMAIN: The RID domain mediates the interaction with nuclear
 CC receptors (By similarity).
 CC -1- DOMAIN: The SPOC domain, which mediates the interaction with
 CC NCOR2, is essential for the repressive activity.
 CC -1- SIMILARITY: Belongs to the Spen family.
 CC -1- SIMILARITY: Contains 1 RID (receptor interacting) domain.
 CC -1- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
 CC -1- SIMILARITY: Contains 1 SPOC domain.
 CC -1- CAUTION: Ref.2 sequences differ from that shown due to erroneous
 CC gene model prediction.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AF356524; AAK52750.1; -
 EMBL; AL034555; CAB85442.1; ALT_SEQ.
 EMBL; AL034555; CAB85444.1; ALT_SEQ.
 EMBL; AL450998; -; NOT_ANNOTATED_CDS.
 EMBL; AL036858; CAB51072.1; ALT_INIT.
 EMBL; AK000882; BAA91405.1; ALT_INIT.
 EMBL; AK022949; BAB14324.1; ALT_INIT.
 EMBL; AB023146; BAA76773.1; -
 InterPro; IPR000504; RNA_rec_mot.
 PDB; 1OW1; 19-AUG-03.
 Pfam; PF00076; RRM; 4.
 SMART; SM00360; RRM; 4.
 PROSITE; PS50102; RRM; 4.
 PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
 PROSITE; PS50917; SPOC; 1.
 Transcription regulation; Repressor; Nuclear protein; DNA-binding;
 Repeat; RNA-binding; Coiled coil; 3D-structure; Polymorphism.
 DOMAIN 1 573 DNA-BINDING (RRM) 1.
 DOMAIN 6 81 RNA-BINDING (RRM) 2.
 DOMAIN 335 415 RNA-BINDING (RRM) 2.
 DOMAIN 438 513 RNA-BINDING (RRM) 3.

1973 SSPKSGSEPRPLPPVNSG-----HATIAITPAK-----NLAPHAS----- 2008
 1513 SSDPHFEPGPMVRGVGTFRDSAGVSPFPKRRPRPKPELLQESLPFPHSSGFLGS 1572
 2009 -PDPPAPAGASD-----PH-----REKTSQKPPSIQ-----ELE 2037
 1573 KPGEFGQAEARDGTALPHIWNRLHTATSRKSYRPTSMPEWMBPLSPFEDVAGTEMS 1632
 2038 LRSLYGHSSVSPGVSPVSPSFTKXGL-----PKH-----LEELDKSHLEG 2084
 1633 QSDGVDLSGDSVSSGPCSORSP-----DGLKGAAEGPKPKPGSGSPNNAVPCGPPG 1688
 2085 ELRKPQKQPVKLGCEAAHLPLRLPESQPSSSPLLQTAQVKVGHQVWVTLAQHISEVIT 2144
 1689 SEPPRRPPAPHDGRKELPREQLP-----PGPIGT-----SQR-----T 1726
 2145 QDTRHHPQQLSAPLAPLYFPGASCCEVLDR-----PPS-DL 2183
 1727 DRGTEPGPIRPS-HRPGPPVQF-GTSDKXDSLRLVWGDLSLKAKELTASVTEALPVSRDW 1784
 2184 YLPPDPHAPAGSPHS-----EGGKSPSPN-----KTSVLGGGEGDI- 2222
 1785 ELLP-----SAAASAEPQSKNLSGHCVPSPSSGQRLYDEVYFGSAGSSQISGSHGLS 1841
 2223 -----EPVSPPEGWTEPHGHSRAV-YPLLYRDGEOTEPEPRMGSKS 2261
 1842 ITSQWRLRPPTSLHPYRSQPLXLPAGPAPPSALLSGVALKGQFLDFTMQATELGKLP 1901
 2262 PGNSTQPPAFF-----SKTENSAMVKSQKEINKLTHNREPEYNISQPGTEI 2313
 1902 AGCVLYPPPSFLYSPAFCSPLDFTSLQVRQ-----DLFSP-SDF 1941
 2314 FNPALITG--TGLMYTYSQAEHASTNGLAIIRKALMGKYDOWERSPPLSANAFNPL 2371
 1942 YSTFLOQGGQSGFLSPGAPAQ-----MLLPWDSQLPVNFGSLPPAPP-----PAPPPL 1992
 2372 NASASLPAAMPITADGRSDHTLTPSGGGKAKVSGRPSRKAKSPAGLSDGDRPPSVS 2431
 1993 SLLPVGALQPPSLA-----VRPPAPATRVLPSPA-----RPPPAS 2029
 2432 SVHSE-----GDCNRRTPLTNRVWEDRPPSAGSTPFPNPLI--M 2469
 2030 LGSRAELHVELKFDQYQKLSSNLGGPSRTPPTGCRSFGSLNRLKATPSTYSYGVFTQ 2089
 2470 RLQAGVWASP-----PPFGLPAGCPLAGPHAWDEEP 2502
 2090 RVDLYQASPPDALRWIPKFWERTGPPREGPSRR-AEEP 2128
 OSA DROME STANDARD; PRT: 2716 AA.
 AC Q81N94: O61603: Q9VEG7;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Trithorax group protein OSA (Eyelid protein).
 GN OSA OR ELD OR CG7467.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RN SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
 RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RX MEDLINE=97415319; PubMed=9271118;
 RA Treisman J.E., Luk A., Rubin G.M., Heberlein U.;
 RT "eyelid antagonizes wingless signaling during Drosophila development
 and has homology to the Bright family of DNA-binding proteins.";
 RL Genes Dev. 11:1949-1962 (1997).
 [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazell R.G., Nelson C.R., Miklos G.L.G.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Beasley E.M.,
 RA Abil J.F., Agbayani A., An H.-J., Bayraktaroglu L., Beasley E.M.,
 RA Ballew R.M., Basu A., Baxendale J., Berman B.P., Bhandari D., Bolshakov S.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brottier P.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-F., Wasarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 [3]
 RN FUNCTION, AND DEVELOPMENTAL STAGE.
 RP MEDLINE=99112962; PubMed=9895321;
 RX Vazquez M., Moore L., Kennison J.A.;
 RA "The trithorax group gene osa encodes an ARID-domain protein that
 genetically interacts with the brahma chromatin-remodeling factor to
 regulate transcription.";
 RL Development 126:733-742 (1999).
 [4]
 RN DNA-BINDING, AND IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM AND SNR1.
 RP MEDLINE=20069333; PubMed=10601025;
 RX Collins R.T., Furukawa T., Tanese N., Treisman J.E.;
 RA "Osa associates with the Brahma chromatin remodeling complex and
 promotes the activation of some target genes.";
 RL EMBO J. 18:7029-7040 (1999).
 [5]
 RN FUNCTION.
 RP MEDLINE=99403006; PubMed=10471712;
 RX Staehling-Hampton K., Ciampa P.J., Brook A., Dyson N.;
 RA "A genetic screen for modifiers of E2F in Drosophila melanogaster.";
 RL Genetics 153:275-287 (1999).
 [6]
 RN IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM, OSA, MOR, SNR1, DALAO;
 RP BAP55; BAP60 AND BAP47, AND FUNCTION AS COACTIVATOR.
 RX MEDLINE=20270023; PubMed=10809665;
 RA Kal A.J., Mahmoudi T., Zak N.B., Verrijzer C.P.;
 RT "The Drosophila brahma complex is an essential coactivator for the
 trithorax group protein zeste.";
 RL Genes Dev. 14:1058-1071 (2000).
 [7]
 RN FUNCTION AS A COREPRESSOR.
 RP MEDLINE=20573925; PubMed=11124806;

2343 EAIIRKALMGKYDWEESPLSANAFLNLSAASLPAAMPITAADGRSDHDTLTSPGGGK 2402
1527 -----WGAPP--RGAAPPAGHPPIQOP--AGVAQWDQHYPPQQGPP 1568
2403 -----AKVGRPSRKAKSPA-----PGLASGD-----RPS 2429
1569 PPPQQQQPPQQQQPPYQVAGPPGQPPQAPQOWAMPQGTAAOSGIAPPGSLRPPS 1628
2430 VSVHSEGCNERTPLTNRVWEDRPSASGTFFP-----YNPLMRILQAGVM--- 2476
1629 -----GPGQONRWGHPAQ--QQSQQQQGVPPPPQQAASHGVSPGLPQVPGMGVMPK 1681
2477 --ASPPPP-----GLPAGSGPLAGPHAMDEPKPL 2505
1682 PYAMPPPPSQGVGQVGGQF-----PGMMMSQKPPPM 1713

RESULT 10
MAPA RAT
ID MAPA RAT STANDARD; PRT; 2774 AA.
AC P34926;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1A (MAP 1A) [Contains: MAP1 light chain LC2]
GN MAP1A.
OS Rattus norvegicus (Rat).
OC Sauriata; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92355629; PubMed=1379599;
RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in one messenger RNA.";
RL J. Biol. Chem. 267:16561-16566(1992).
CC -!- FUNCTION: Structural protein involved in the filamentous cross-bridging between microtubules and other skeletal elements.
CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate with MAP1A and MAP1B proteins.
CC -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE THEIR MORPHOLOGY.
CC -!- DOMAIN: The basic region containing the repeats may be responsible for the binding of MAP1A to microtubules.
CC -!- PTM: Various serine residues may be phosphorylated by CAMP kinase.
CC -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH BOTH MAP1A AND MAP1B.
CC -!- SIMILARITY: TO MAP1B.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

EMBL; M83196; A848069.1; -;
PIR; A43359; A43359.
Microtubule; Repeat; Phosphorylation.
CHAIN ?2465 2774 MAP1 LIGHT CHAIN LC2.
FT CHAIN 2465 2774 LYS-RICH (BASIC).
FT DOMAIN 309 496 11 X 3 AA REPEATS OF K-K-[DE].
FT DOMAIN 336 541 1.
FT REPEAT 336 338 2.
FT REPEAT 415 417 3.
FT REPEAT 420 422 4.
FT REPEAT 424 426 4.

1353 GSITQIGIPRSYEAQEDYLREAKLLKREGTPPPPPSRLDTLAYKTQALGPLKLKPAHE 1412
730 G-----GAPGGAMGNH-----VQKGTPTPPVVG-----GP-----PPQ 760
1413 GLVATVEAGRSIHIEIPRELRHTPELPLAPRPLKEGSIOTGTPKYDTGASTGSKKD 1472
761 G-----SGSPRLNLYL-KOHLQHGKGGYGGSTP-----PQG-PQGVNG--PTGM----- 801
1473 VRSLGSGRTFFPPVH-----PLDVMADARALERACYEELSKSRREGTASSG 1519
802 -----HPCMPGPHMGPHGPTNMGPPTSTPTPQSMQLQ-----GGQPGQCGASG 847
1520 GSIARGAPVIVPELKGKRPQSLTYEDHGAPAGHLPKRGSPVTMREPTPRLOEG--SLSS 1577
848 G-----PESGPEH-----ISQDNGISSG--PTGA-AGMHAVTSVVTTGPDGTSMD 891
1578 KASQDRKLTSTPREIAKSPHSTVPEHHHPHPISPYEHLLRGVGVLYRSHRPLAFDPTSI 1637
892 EVSQOSTLSNAAASAGEDPQCTTKRKN-----DP----- 922
1638 PRGIPLDAAAYLPHRLAPNPTYPH-----LYP-----PYLIRGYPTAALENRQTIND 1688
923 -----YSQSLAPSTSPHVVMPHGGGPGGEGYDMSSPPNPNRAGSPQVFNH 970
1689 YITSQMHNTATAMAQADMLRGLSPRESSLALNYAAGPRGID-----LSQVHLPLV 1744
971 VVPQEPFRSTITT-TKSDSLCKLYEMDN-----PDRGWLDKLRAFMEERTPTITA 1023
1745 PTPGTPATMDRLAYLPTAPQPPSSRSHSSPLSPGPTHTLTKPT-----TSS 1793
1024 CPTISKQPLDLRL-YIYVKE-----GGFVETKSKTDIAGLLIGASS 1069
1794 SERERDRDRDREREKSIILT-----STTTVEHAPIWR-----POTEGSSGSGS--SGG 1843
1070 SAAITLR-----KHYTKNLTTFECHFDRGDIDPLPIIQVEAGSKKTKAKAASVPSPG 1122
1844 GGGSSSRPASHGA-HQHSPIGPTQDALQRPVSLVHNTGMK-----LIITAVE 1891
1123 GG-----HLDAGTNTSGSNSQDSFPAPPGSAPNAIDGYPGYGGGPPYVVASGQ 1174
1892 BSKPTVLRSTSSVVRPAATPPATHCPGLGTLGVPTLMPEVLLPKPEAPRVARPERP 1951
1175 PYVATAGQORPPSQNNPQTPHGAATAAAG--DNI-----SVSNPFEDP-IAAGGP 1225
1952 RADTGHAFLLAKPPARGLEPASPP-----SKGSEPRPLVPVSGHATIAITPA--KN 2001
1226 GSQGT-----PGQGGPGGAASGAGAVGAGVGPGQPPHPPPHSPHTAAQQAAGQHQ 1279
2002 LAPHASPPPPASADPHREKTSKPSIQEELSLGVLHSGSYSP-----GVEPVS 2057
1280 QHPQHQHGLGFPFPQQQGGQGGQPP-----SVG--GGPPAPQQRHGPQVPPS 1329
2058 PVSSPSLTHDKGLPKHLELDKSHLEGELRPQKPGVPLGSGEAAHLPHLRPLPESQPS 2117
1330 P-----QHVRAAGAVYPPGSG-----YPTPVSRTPGS 1359
2118 PLLQTPAGVKGHORVTLAQHISEVITQDTRHHPOQLSAPLPAPLYSPFGASCPVLDLR 2177
1360 P-YPSQAGY-----QYSGSDQYNATGPPQPGQPGPG-----Q 1393
2178 RPPSDLYL-PP--PDHGAPARGS-----PHSEGGKRSRPPNKTSLVGGGEDIEPV 2225
1394 YPPQNRNMYPPYGEAEAPTGANQVGYGSRYSQPPPGPQPPPTQTVAGGPPAGGAPG 1453
2226 SPPEGMTPEGHRSASVYLLYRDEQTEPRSMGSKPGNTSQPPAPFSKLKTESNANVKS 2285
1454 APP-----SSAYP-----TGRPSQDYIYPPPPQSQPPQRHHPDFIKD 1490
2286 KQEQINKLNTNNEFEYNI SQGTETIFNMALTG---TGLMYSQAVQEHASTMGL 2342
1491 -----SQPYFGYNA-----RPQIYGAWQSGTQYRPPYSSPAPQN--- 1526

FT	REPEAT	427	429	5.	
FT	REPEAT	431	433	6.	
FT	REPEAT	436	438	7.	
FT	REPEAT	440	442	8.	
FT	REPEAT	444	446	9.	
FT	REPEAT	449	451	10.	
FT	REPEAT	539	541	11.	
SQL	SEQUENCE	2774	AA; 239526	MM; 3DEFT4427BA9D7D7	CRC64;

Query Match

Best Local Similarity

Matches

3.7%; Score 492.5; DB 1; Length 2774;

20.5%; Pred. No. 7.2e-09;

553; Conservative 278; Mismatches 1002; Indels 861; Gaps 132;

Qy	77	ERSQELHURPSHSYLPBLGKSEMEFISKRPRLLELLDPDILLRPSPLLATQOPAGESDLT	136
Db	398	DKAGGHLK-EKISKLEE--KKDKKKIKKEREKELKE-----EGRKEEK	440
Qy	137	KDRSLTGLKLPVSPSPHTDPELEVPRLSK---BELLQNMDRVDRREITWVEQQISKL	193
Db	441	KDAKKDKRKTKEVVKLSKPDILKPTTPEVRKTLTKAKAGRVKVKDK-----GRA	491
Qy	194	KKKQQQLEEEAAKPEPEKPVSPPIESKHSLV-----QIYDENRKKAAEAHRILE--	246
Db	492	ARGEKELSSERTPP-AQKGAAPAAVSGHRELALSSPEDLTQDFEELKREERGLLAER	550
Qy	247	--GLGQVELPLYNQPSD-TRQYHNTKINQAMKGLILYKFRNHARKQWKQPCORYD	303
Db	551	DTGLG---EKL---PADATEQGHPSAAIQVTPSGPVLE-----584	
Qy	304	QLMEALEKQVRIENNP-----RRRAKESKVREYVEKQFPIRKORELOERMQSRVQGR	358
Db	585	--GEHVERKEVVPDSDGDKGKTWRGPDSDGAEVKEKEKTWEERKQREAB-----631	
Qy	359	SGLSMSAAR--SEHEVSEIIDLSEQENLEKMRQLAVIPMLYDADQORIKFINMGLM	416
Db	632	LGPENTAAREBSEAEVKEDVIEKAELEMEB-----THPSD-----667	
Qy	417	ADPMKVYKDRQVMNMWSEQEKETREKPMQHPKPNGLIASPLEKTVAEVCVLYYLTKKN	476
Db	668	-----EGEETKAESFYQKHTQELKASPKSR-----694	
Qy	477	ENYKSLVRRSYRRRGKSQQQQQQQ-----QQQQQQQQQPMPSRSEQEKDEKEK	525
Db	695	---HALGGRDLGFGKAPKEKETAFLSLATPAGATEHVSYIQETIPGYSETQTTISDE	751
Qy	526	EKEAEKEEE--KPEVNDKEDLLKKTDDTSDGEDNDEKBAVAKGRKTANSQGRR---K	579
Db	752	EIHDEPDERPAPRFPFTSTYDLSGPEGPPEASQAADSAVPASSSKTYGAPETELTVPP	811
Qy	580	GRITSRMANEANSSEALTPQOSALAM-----ELNESSHTWEEMETAKKGLLEHCRN	633
Db	812	NMVAAPLAEHEHVSSAYSITCEDKLSFATSVAESDQSVASLUTAPOTETGKSSLL----	866
Qy	634	WSAITARVMGSKTVSQCNFYNFKKRNQLDBEILQQHKLKMEKERNARRKKKAPAAASEE	693
Db	867	LDTVTSIPSSRT-----BATQGLDIVPSAGTI-----STSSLEE	901
Qy	694	---AAFPVVEDEMEASGVSGNEEEMVVEAEALHASNEVPRGECGSPATVNNSSDTE	750
Db	902	DKGPKSP-----CEDFSVTGESEK-----KGETVGRG-LSGEAKVAGEEKVY-	943
Qy	751	IPSHTTAAKXDTGONGPKPATILCAGDP--PPGPPTPPRRTSRAPITPTPASEATGATPP	808
Db	944	-----VTISEKLSGQYA----AVFGADGHTLPPGPALGEVEERC-LSPDDSTVQKASGP	992
Qy	809	PPAPPPSAPP-----PVVPKEEKE-----ETAAAPPV--EGBEQK-PPAAEEELAVD	854
Db	993	PSGPPSAAHTPFHQSPVEDKSEPRDFQEDSWGETHSPGVSKEDESRQTVKPGPEGTSE	1052
Qy	855	TGKAEPKPVSECTEEAEEGPAKGDAAEAATAB--GALKAEKKEGGSG-----R	902
Db	1053	EGKG--PPTRSPQADMVPSIAGGTGCTCIQLLPEQDKAIVPETGEAGSNLGAFTLPGEVR	1111

QY	903	ATTAKSGAFQDSSATCSADEYDEAGGDKNRLLSPRFSLLTPTGDPNANASPOKPLD	963
DB	1112	TSTEATEPQKDEVLRFTDQSLSPEDAESLVLSVSPDIT-----KQETPRSPCS	1163
QY	963	LKQLKQRAAIPPTQVTKVHPREDAAPTKPAPPPONLOPESDAFQOQSSSRGK	1022
DB	1164	LKEQOPHKDWPVMSPEDTQSLSFSESPSKET-SLDSISKQJLSPESLQTOFQELNLGK	1222
QY	1023	-SRSPAPPADKEAFABEAQKLPDGPWCWTSGLPFPVPPREVIKAS-----PHAPDP-SAFS	1076
DB	1223	EERGFVMKAE-----DDSCHLAPVSIPEPHRAIVTSPSTDETAGTILPGGSFS	1269
QY	1077	YA-----PPGHPLPLGUH--DTARPVLPRPTTISNP-----PP	1107
DB	1270	HSALSVDRKHSPGEITPGGHFTSDSSLTKSPESLSSPAMEDLAVEWEGKAPGKEKEPE	1329
QY	1108	LISAKHP--SVLEROIGALISQMSV-----QLHVPYSEHAKAP-----	1144
DB	1330	LKSTROOKQOILPEKVAVVVEODLIHQDGALENKPKRQOQDKTEQKGRDLDEKDTA	1389
QY	1145	---VGPVWTGMLPMDPKKLAPGSGVKQQLSPRGO-AGPPESLGVPTAQEASVLRGTA	1199
DB	1390	AELDKGP-----EPKE-----KDLDRDQOQKAGPPAE-----KDKASEQRDTD	1428
QY	1200	LGSVPGGSI7KIGPSTRVPSDSAITYRGS-----ITHGTADVLTKTITRIIGED-	1250
DB	1429	LQOT-----QATEPRDRAQERDSEKDKSLERDRTPEE---KORI--LVQEDR	1473
QY	1251	-----SPRDLR-----GRESLPGKHVIYE-----GKKGHVLSEYG--GM	1284
DB	1474	APEHSIPEPTQTRAPDRDKGTDDEQKEEASEEKEQVLEQKOWALGKEGTLTQEAATAE	1533
QY	1285	SVTQCSKED--GRSSSGPPHETAAPKRTY-----DWMSEGRVGAIRSSASIEGLMG	1332
DB	1534	QKDETLKEDKTQOGKSSFVEDKTTTSKETVLQDQSAEKADSVQDGAALKEKTRALG--	1591
QY	1333	RAIPPE-----RHSPHLKQHHIRG---SITQIGIPRSVYEAQEDYLRREAKLLKREGT	1383
DB	1592	EESPAEGSKAREQEKYKWEQDVVQWRETSPTRGEP--VGQKE-----PVPAMEGK	1642
QY	1384	PPPPP-----PSRDLT---EAYKTOALGPLKLKAHE-----GLVATVKEAGRS-IH	1430
DB	1643	SPEQEVRYWDRDITLQODAYWRELSCDRKVWPFPHELDGGGARPRYCEERESTFLDEGPD	1702
QY	1431	EE---LHRHTPELPLAB--RPLKEGSI7QCTQPLKYDGTGASTTCSKKHDVRSLLTSGPRTF	1484
DB	1703	EQEITPLQHTPRFWSFTDQFOQEPLPOKGLEVERWLAESPVLGPPPEEDKLTSPFEII	1762
QY	1485	PPVHPLDVNMADARALACYEES-----LKSRPGTAS-----S	1517
DB	1763	SPPASPPEM--TGQVPSAPQSSPVPDTESTAPMRNEPTTPSWLAEIPWVPKDRPLPP	1820
QY	1518	SGGSIARGAPVIVPELGKPROSLTY-----EDHGAPPAGHLPRGSPV	1560
DB	1821	APLSPAPAPTPPAPEHTP--VPFWSGMLAEYDVVAOVEGAAELEGPPYS---PLGKOY	1875
QY	1561	TMRPTPLQEGSLSSSKASQDKLSTPREIA-KSPHSTVPSHHPHIPISPYEHLLRGVS	1619
DB	1876	RKAERGEKGAGADPDSSFSKVPFAGESLATRDTEQTEPQREPTVPDB---RSFQ	1932
QY	1620	GVOLYRSHIPLAFDPTSIPRGIPLDAAAAAYLPHILA-----PNPTYPH	1663
DB	1933	YADIYEQMMLTGLGPACPTREPFLGASGDW--PPLHSTKEAAGCNTSABKETSASPAPQ	1990
QY	1664	-----LYPPYLIRGVPDTA-----ALENQTIINDYITSQOMH	1696
DB	1991	NLQSDTPAFSYASLAGAVPP---ROEPDGPVNVFSPITPPAVPPRAPISLSKDLSPPLN	2047
QY	1697	HNTATAMAQR-----ADMRLGL-----SPRE-----	1717
DB	2048	GSTVSCSPDRRTSPKETGRGHWDGTDNSDLSEKGAEREQEKETRSPSPHHMPMGHSSL	2107
QY	1718	-----SSL-----ALNYAAGRGGIIDLQVPHLPVLVLPPTQGTATAM-----	1755

```
Db 2108 WPETEAYSSLDSSHLGSRVPSLDPPASAFSSQLQAP--PQL--PSPAPRSPAPCGSL 2163
Qy 1756 ---DR-LAYLPTAPQPFSSRHSS-----SPLSPGGP--THLTXTPT 1790
Db 2164 AFGDRALALVPGTPT--RTHDYLEVTKAPSLDSSILPQLPSPSPGGPLSLNLPAS 2221
Qy 1791 TSSSERDRDRDRDREREKSLTSTTTVEHAPIWRPTEQSSGSSGGGGGSSSR 1850
Db 2222 FALSE-----GSSSEATTPTVSSVAERFP--PGLE-----AAEQSAEGLGSGKE 2263
Qy 1851 PASHSIAHQHPISPRITODALQORPSVLNHTGMKGIITAVEPSKPTVLRSTSTSSVPA 1910
Db 2264 SAAHS-LWDLTFLSPASADLAPA-----PA 2290
Qy 1911 ATFFPATHCPGLGTLGDVYPTLME-----PVLLPK-----EAPR 1945
Db 2291 PAPAPAGLP-GDLGDGTLPCRPCTGELTKKPFLLSPGDHEANGPGETSLNPPGFVT 2349
Qy 1946 ARPERPRADTGHAF-----LAKPPARSGLPEPASPSKSGSEPRPLVPVSGHATI 1994
Db 2350 ATAEEEAEPHAWERGSWPEGAERSRPDTLLSSEQPLRPGKSGG-----PPCSSLSEV 2405
Qy 1995 ARTP---AKNLAPH--HASP---DPPAPPASADPHREKTSQKPSIQLEL-----RS 2040
Db 2406 EAGPGCATDPRPHCGELSPFLNPLPFS-----TDDSLSTEARLAGKGGRR 2456
Qy 2041 LQYHGSSYSP-----EGVFPVSPVSPSLTHDKGLPHLEELDKSHLEGLRKPQG---P 2093
Db 2457 VGRPATGCPMADETPTTSASDSGSSQSDSDVPETEECPSTAEAAALSDSDGDFLP 2516
Qy 2094 V-KLGG-----EAAHLPHRLPESQPSPLLOTAPGVKGHORVTLAQHISEVITQ 2145
Db 2517 VDKAGVSGTHHPRPGHDPPTPLDPRPSP-----RP 2550
Qy 2146 DYTRHHPPQSLAPLAPLYSPFGACPVLDLRRPSPDLVLPDPHGAPARGSPHSEG--- 2202
Db 2551 DVMCMADPEGLSS-----ESGRVERLEKCR-----PGRAPGRAPKSPARRL 2593
Qy 2203 ---GKRSPFNKTSVLGGEDGIEFVSPPEGMT-----BFGH---SRSAVYPLLVRDGGQT 2252
Db 2594 DIRGKSPFTFGKPVDRSRTVPRSTPSQVTSAAEKDGHSPMSKGLV-----NGLKA 2647
Qy 2253 EPSRMGSKPGNTSQPPAPFASKLTESNANWKSKEINKLNTH-----NRNEP 2302
Db 2648 GSTALGSK---GGSGPPVYDLAYTPNHCSGKTDQDFRVRASYVYVVGNDP 2698

RESULT 11
MLL2_HUMAN
ID MLL2_HUMAN STANDARD; PRT; 5262 AA.
AC O14686; O14687;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related protein).
DE protein).
GN MLL2 OR ALL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX MEDLINE=97388474; PubMed=9247308;
RA Prasad R.; Zhadanov A.B.; Sedkov Y.; Bullrich F.; Druck T.;
RA Rallapalli R.; Yano T.; Alder H.; Croce C.M.; Huebner K.; Mazo A.;
RA Canaan E.;
RT "Structure and expression pattern of human ALL1, a novel gene with
RT strong homology to ALL1-1 involved in acute leukemia and to Drosophila
RT trithorax."
RL Oncogene 15:549-560(1997).
RN [2]
```

```
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=22371456; PubMed=12482968;
RA Goo Y.-H.; Sohn Y.C.; Kim D.-H.; Kim S.-W.; Kang M.-J.; Jung D.-J.;
RA Kwak E.; Barlev N.A.; Berger S.L.; Chow V.T.; Roeder R.G.;
RA Azorsa D.O.; Meltzer P.S.; Suh P.-G.; Song E.J.; Lee K.-J.; Lee Y.C.;
RA Lee J.W.;
RT "Activating signal cointegrator 2 belongs to a novel steady-state
RT complex that contains a subset of trithorax group proteins.";
RL Mol. Cell. Biol. 23:140-149(2003).
CC -1- FUNCTION: May be involved in transcriptional regulation.
CC -1- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/
CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins
CC MLL2 and MLL3, and ASH2/ASCL2.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Evtent=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=O14686-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O14686-2; Sequence=VSP_008563, VSP_008559;
CC Name=3;
CC IsoId=O14686-3; Sequence=VSP_008560;
CC -1- TISSUE SPECIFICITY: Expressed in most adult tissues, including a
CC variety of hematopoietic cells, with the exception of the liver.
CC -1- MISCELLANEOUS: This gene mapped to a chromosomal region involved
CC in duplications and translocations associated with cancer.
CC -1- SIMILARITY: Belongs to 5 PHD-type zinc fingers.
CC -1- SIMILARITY: Contains 1 post-SET domain.
CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
CC -1- SIMILARITY: Contains 1 SET domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF010403; AAC51734.1; -.
CC EMBL; AF010404; AAC51735.1; -.
CC PIR; T03454; T03454.
CC PIR; T03455; T03455.
CC Genew; HGNC:7133; MLL2.
CC MIN; 602113; -.
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0003700; F:transcription factor activity; TAS.
CC GO; GO:0007048; P:oncogenesis; TAS.
CC GO; GO:0006366; P:transcription from Pol II promoter; TAS.
CC InterPro; IPR003889; Fyric-C.
CC InterPro; IPR003888; Fyric-N.
CC InterPro; IPR000910; HMG_12_box.
CC InterPro; IPR003616; PostSET.
CC InterPro; IPR006118; Recombinase.
CC InterPro; IPR001214; SET.
CC InterPro; IPR001965; Znf_PHD.
CC InterPro; IPR001841; Znf_ring.
CC Pfam; PF00628; PHD; 5.
CC Pfam; PF00856; SET; 1.
CC SMART; SM00542; FYRC; 1.
CC SMART; SM00541; FYRN; 1.
CC SMART; SM00398; HMG; 1.
CC SMART; SM00249; PHD; 7.
CC SMART; SM00508; PostSET; 1.
CC SMART; SM00184; RING; 3.
CC SMART; SM00317; SET; 1.
CC PROSITE; PS00868; POST_SET; 1.
CC PROSITE; PS0280; SET; 1.
CC PROSITE; PS01359; ZF_PHD_1; 5.
CC PROSITE; PS50016; ZF_PHD_2; 5.
```


Db 2110 PRQPPPPSCALLPPRS-LPSDFSRVPVSPQSSQSSPLTPRPLSAEAF- 2161
Qy 1044 GPPCWTSGLPFPVPPREVITKASHPADPSAFSAVPPGHL-PLG-LHDTARPVLPRP 1099
Db 2162 -PSPVTPRF-OSDPYS--RPPSRPQSRDPFAPLH--KPPRPQP 2199
Qy 1100 PTISNPPLTSSAKHPSVLERQICAI SQGMSVQLHV-PYSE-HAKAPVG-----PVT 1149
Db 2200 PEV-----AFKAGSLAHTSLGA--GGFPAALPAGAGELHAKVPSGPPNVRSPGT 2249
Qy 1150 -MGLPLEM-----DPKLAFTSGVKQEQLSP-----RGOAGPESIGVPTAQEAS 1193
Db 2250 CAFVGTSPMRFTFPQAVGESLXPP--VPQGLPPPHGINSFPGPTLGKQSTNYT 2306
Qy 1194 VL-----RGTAIGSVPGGS-ITKGIPISTRVPS-----DSAITYRGSIHTGPPADVLY 1239
Db 2307 VATGNFHPGSGPLGFGSSGTGESYGLPRLPSPVLPAPPADGSLPY--LSHGASQ--R 2360
Qy 1240 KGTITRIIGEDSPRLDRGREDSLPKGHVIEGKGHVLSYEGGMSVTOCKED----- 1293
Db 2361 SGITSPVEKREDPG--TGMGSLATAEL--PGTQDPGMS--GLSOTELEKQOROLR 2412
Qy 1294 -GRSSGPPHETAAPKRTYDMMEGRVGRATSSASIE-----G 1329
Db 2413 ELLIROQIORTLROKETAAAGACVPPSGWGAEPSSPAFEQLSRGOTFACTQDKSS 2472
Qy 1330 LMGRAIPPERHSPHLLKEQHIRGSIQTGIPRSVVEAQEDVLRREAKLLKREGTPPPP-- 1387
Db 2473 LVG-LPPSK-----LSGPILG--PGSF--PSDDLRS-----PPPPAT 2505
Qy 1388 PPSRDL-----TEAVKTOA--LGLKLPKPAHEGL-----VATVKEAGRSIH--EI 1428
Db 2506 PSSMDVNSRQVGGSAFYQRYAPYFSGPLFQQQQQLWQQQQAATAMRPFAMARFST 2565
Qy 1429 PREEL-RHTPELELA-----PRPLKEGSIQTGTPKLYDTGASTTGGKKHDVRLSIGSPG 1481
Db 2566 PGPELGRQALGSLAGISTRLPGP-----GEPVPGPAGPAQFIELRHNVQKGLGPGG 2617
Qy 1482 RTFPVPVHLDMADARALACRYEESLSKSRGCTASSGSGSTARGAPVIVPELGRQSP 1541
Db 2618 TFPF-----TEAVKTOA--LGLKLPKPAHEGL-----VATVKEAGRSIH--EI 1428
Qy 1542 TYEDHGAFFAGHLPRGSPVTMREPTPLQEGSLSSKASQDRKLTSTPREIAKSPHSTVP 1601
Db 2635 SEDPH-----FLAEGUL-----GLAVSGLPPQKPSAPPAP-ELNLSLHT-- 2674
Qy 1602 EHHPHPISP-----YELLRGVSGVDLYRSHIPAFDPTSIPIRGIP-----LDAAAAYYL 1651
Db 2675 ---PHTKGTPLTGLVLRNPPSPSTELGRPN-PLALEAGKLPCEDPELDDDDFDAKHALED 2730
Qy 1652 PRHLAPNTYPHL-YPPVYLIRGYPTAALENRQT----- 1684
Db 2731 DEELA-----HLGLGVDAVAKGDDDELGTLENLTNDPHLDLLNGDEFLLAYTDPDLDT 2784
Qy 1685 -----IINDYITTSQQMHNTATAMAQRAADMLRGLSPRESSIALNVAAGPRGIIDLSQVPH 1739
Db 2785 GDKDIFNEHLKLV-----SANEERERALLGVPE-----GFLG----- 2820
Qy 1740 LPVLVPPPTGTPATMDRLAYLFTAPQPF--SSRHSSSPSLSPGGPHTLTKPTTSSSRE 1797
Db 2821 -PEERPP-PAADASEPRLASVLPEVKPKVEEGGRHPS-----PCQFT----- 2860
Qy 1798 RDRDRDRDREREKSLTSTTTVEHAPI-----WRPG-----TEQSSGSGSGG 1842
Db 2861 -----IATPKVEPAAPAAANSLGLKPKGSMMSGRDTRMTGTFPSSG 2902
Qy 1843 -----GGGSSSRPASHSHAHQHSPLSPRTQDALQORPSVLHNTGMKGIITAVEPSKPTV 1897
Db 2903 HTAKASGATGCPAH-----LLTSPSLSGPGSSSLEKFE----- 2939
Qy 1898 LRSTSTSSVRRPAATPPPATHCPLGGLTLDGVVPTLMEPVLLPKPEAPRVARP--ERPRADT 1955
Db 2940 LESGALTLPGGPAAS-----GDELD-----KWESSLVASSELPLIEDLLEHEKKE- 2984

Qy 1956 GHAFKAPPARSGLEPASSPSKSGSEPRPLVPPVSGHATARTPAKNLPHHAGSPDPAPP 2015
Db 2985 ---LQKKQSLASQLQFAQQQQQQQHSLLP----- 2015
Qy 2016 ASASDPRHREKTSQKPFISIQLERLSLGYHGSSYSPEGVFVPSVSPSLTHDKGL-PKHL 2074
Db 3018 QAMSLPHEGSSPSLAGSOQL---SLGL-AVARQPGLPQLMPTQPPAHALQORLAPSMA 3073
Qy 2075 EELDQKH-LEGEURPKQPGVKLGGEAAHLPHLRPLPESOP---SSPLLQTAQVKGHOR 2131
Db 3074 MVSNQGHMLSGQ-----HGGQAG---LVPOQSSQFVLSQKPMGTMPMSMCKPQ 3119
Qy 2132 VVTLAQHISEVI--TQYTRHHQQLSAPL-PAPLYSFPGAS-----CPVLDLRR 2178
Db 3120 QLAQQOQLANSFPDDTDLKFAEDIIIGPIAKAKVWALKIKKVMAGSGIVAFGMNRQ 3179
Qy 2179 ---PPSDLYLPPPDHGAARGSPHSGEGRKSRPEPNKTSVILGGEGDIEVSP 2227
Db 3180 VSLLAQRLSGPSSDL---QNHVAGSGQERSAGDPSQPRPNPPTFAQGVINEADQRY 3235
Qy 2228 PEGMTEPHGSRSAVYPLL-----YRGEQTEPSRMGSKSPGNTSQPPAPFSKLTES 2278
Db 3236 EEWLF---HTQQLLMQLKLEEQIGVHRKSRKALCAKORTAKKAGREFPEADAEL--- 3289
Qy 2279 NSAMVKSQKQIINKLNTNRNEPEYNI SQPTEIFNMPAITGTGLMT-YRQAVQSHAS 2337
Db 3290 --KLVTQQSKIQKQLQVVRKQKEH-----TNLMAEYRNKQQQQ--- 3328
Qy 2338 TNMGLEAIIRKALMGKYDQWESPPLSANAFNPINLASASIPAAAMPITAAAGRSHTTSP 2397
Db 3329 -----QQQQQQQQHSAVLASP-SQSPRLTLKPLGQLLPG---HGLQPP 3369
Qy 2398 GGGKAKVGRPSRKAAPGLASGDRPPSVSVHSEGDGNCNRTPLTNRVWEDRPSA 2457
Db 3370 QG-----PPGQAGG----- 3379
Qy 2458 GSTFPFYNLMRLQAGWASPPPPGLP-----AGSGPLAGP 2494
Db 3380 -----LRLTPGMALPGPGGFFLNTALAQQQQQQSHGAGSLAGP 3420

RESULT 12

SPEN DROME
ID SPEN DROME STANDARD; PRT: 5560 AA.
AC Q8SX83; Q9NH17; Q9U6C3; Q9VPL1; Q9VPL2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Split ends protein.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4), FUNCTION, ALTERNATIVE PROMOTER
RP USAGE, TISSUE SPECIFICITY, DEVELOPMENTAL STAGE, AND MUTANTS E9 AND
RP D57.
RC TISSUE=Embryo;
RX MEDLINE=20025936; PubMed=10556062;
RA Wiellerte E.L., Harding K.W., Mace K.A., Ronshaugen M.R., Wang F.Y.,
RA McGinnis W.;
RT "spen encodes an RNP motif protein that interacts with Hox pathways
RT to repress the development of head-like sclerites in the Drosophila
RT trunk."
RL Development 126:5373-5385 (1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Embryo;
RX MEDLINE=20157049; PubMed=10655223;
RA Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Laverty T.,

RA Suh C., Voas M., Williams A., Rubin G.M.;
RT "A genetic screen for novel components of the Ras/mitogen-activated
RT protein kinase signaling pathway that interact with the yan gene of
RT Drosophila identifies split ends, a new RNA recognition motif-
RT containing protein";
RL Genetics 154:695-712(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1), SUBCELLULAR LOCATION, AND TISSUE
SPECIFICITY.
RT TISSUE=Embryo;
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10704397;
RA Huang B.H., Wu S.C.-Y., Shin Y.-A., Luo L., Kolodziej P.A.;
RT "split ends encodes large nuclear proteins that regulate neuronal
RT cell fate and axon extension in the Drosophila embryo";
RL Development 127:1517-1529(2000).
RN [4]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke S., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Ducan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhou L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [5]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.-Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [6]
RP SEQUENCE OF 424-2002 FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;

RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [7]
RP FUNCTION.
RX MEDLINE=20253107; PubMed=10790398;
RA Lane M.E., Blend M., Heidmann D., Herr A., Marzodko S., Herzig A.,
RA Lehner C.F.;
RT "A screen for modifiers of cyclin E function in Drosophila
RT melanogaster identifies Cdk2 mutations, revealing the insignificance
RT of putative phosphorylation sites in Cdk2";
RL Genetics 155:233-244(2000).
RN [8]
RP FUNCTION ON EGF RECEPTOR PATHWAY.
RX MEDLINE=20414403; PubMed=10959845;
RA Chen P., Rebay I.;
RT "split ends, a new component of the Drosophila EGF receptor pathway,
RT regulates development of midline glial cells";
RL Curr. Biol. 10:943-946(2000).
RN [9]
RP FUNCTION ON WG PATHWAY.
RX MEDLINE=22668876; PubMed=12783785;
RA Lin H.V., Doroquez D.B., Cho S., Chen F., Rebay I., Cadigan K.M.;
RT "Split ends is a tissue/promoter specific regulator of Wingless
RT signaling";
RL Development 130:3125-3135(2003).
RN [10]
RP FUNCTION: Probable corepressor protein, which regulates different
RP key pathways such as the EGF receptor and Wg pathways. Involved in
RP neuronal cell fate, survival and axon guidance, cell cycle
RP regulation and repression of head identity in the embryonic trunk.
RP May act with the Hox gene Deformed and the EGF receptor signaling
RP pathway. Positive regulator of the Wg pathway in larval tissues
RP but not in embryonic tissues. May act as a transcriptional
RP corepressor protein, which repress transcription via the
RP recruitment of large complexes containing histone deacetylase
RP proteins.
RN [11]
RP SUBCELLULAR LOCATION: Nuclear.
RN [12]
RP ALTERNATIVE PRODUCTS:
RX Event=Alternative promoter;
RX Comment=2 isoforms, 1 (shown here) and 2, are produced by use of
RX alternative promoters;
RX Event=Alternative splicing; Named isoforms=4;
RN Name=1;
RX IsoId=Q8SX83-1; Sequence=displayed;
RN Name=2;
RX IsoId=Q8SX83-2; Sequence=VSP_008565, VSP_008566;
RX Note=No experimental confirmation available;
RN Name=3; Synonyms=SpentL;
RX IsoId=Q8SX83-3; Sequence=VSP_008567;
RX Note=Produced by alternative splicing of isoform 1;
RN Name=4; Synonyms=Spens;
RX IsoId=Q8SX83-4; Sequence=VSP_008565, VSP_008566, VSP_008567;
RX Note=Produced by alternative splicing of isoform 2;
RN [13]
RP TISSUE SPECIFICITY: Ubiquitous. Expressed prior to cellularization
RP in stage 3 embryos, and in blastoderm cells, including pole cells.
RP Expressed throughout the rest of embryogenesis. Later, it is
RP expressed at higher level in epidermal cells and CNS.
RN [14]
RP DEVELOPMENTAL STAGE: Isoform 3 is expressed both maternally and
RP zygotically.
RN [15]
RP SIMILARITY: Belongs to the Spn family.
RN [16]
RP SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
RN [17]
RP SIMILARITY: Contains 1 SPOC domain.
RN [18]
RP CAUTION: It is uncertain whether Met-1 or Met-7 is the initiator.
RN [19]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (see <http://www.ebi.ac.uk/ebis-sib.ch/announce/>)

Qy	1326	SIEGLMGRATPPERHSPHLLKQBHHIRGSGITQGIIPRSYV-----	1366
Db	3879	-----QHNOQHLNQJLHAQOHPPTKQHQAOQOQFNQOIQHQSQOQHQVQOQN	3926
Qy	1365	EAQSDYLRREAKLLKRGSTPPPPPSRDLTEAYKTOAL-GPLKLPKPAHEGLVATVKEAGR	1423
Db	3927	QAQOQHLSQQHQSQQQQNLQOQHQAQQOQLQIQKLQOMHGFPQQQKSPQG-----	3982
Qy	1424	SIHIEPREELHTPELPL-APRPLKEGSGITQGTPLKYDYTCASIT-----	1466
Db	3983	STSFASQO--HNSQLPARGVPOQHQPQQLSHSSPCKNLTIVSNQGVQPAILTRVGSH	4040
Qy	1467	-----GSKKHVRSLLIGSGRTPFPVPHLDVMDADARAL--ERACVEESLKSUR	1511
Db	4041	SQPNQOQQLPHQOQSSGHPHQQLSSPCANLPLQTPLNVINTQNTPKIIVOHIAQNQVPP	4100
Qy	1512	PGTASSGGSGTARGAPVIVPE-LGKRPQSPLTYEDHGAPFAGHLPGRSPVTMEPTPLQ	1570
Db	4101	PQT-----QGNALHYQNGQKDSPPP-----GHV-----EPTP-----	4128
Qy	1571	EGSLSSKASQORKLTSTPREI-----AKSPHSTVPEHHHPIS--PYEHLRLGVS--	1619
Db	4129	--AMSAQKTSSESVSIVRTPPTTGLAVISANTVCSLLTEENLIKISQPKODELIEQDSKE	4186
Qy	1620	-----GVD--LYRSHIPLAFDPTSPRGIPLDAAAYVLPRLHAPNTPHYPLYP	1666
Db	4187	VDSYWSAKEVNIIDSVIKLDTPLA--SKDAKRAVEMQAITAPIPN---POPQNSQMAQ	4241
Qy	1667	PYLLRGVYDPTAALENROTIIINDYITLSQOMHHNTATAMAQRA-----DMLR	1711
Db	4242	ETAL--PTTSMVSNNNS---NDHDTED-----TETROLPPAKPPIPTVGRPPQGRGSAKR	4291
Qy	1712	GLSPRES-----SLALN--YAAGPRGIIIDLSQVP-----HLPVLVPPTPGTPA-----	1752
Db	4292	GRQPRGAKKVGFFLNSVTAAPGVDLSLVQPGDNGVOTRLRKPVTAPVTRGKGRPPRN	4351
Qy	1753	-----TAMDRLAYLPTAPQPFSSRRHSSGPLSGGPTHILTKPTTTSSSRERDRDRERD	1805
Db	4352	LLLQOQQLQOQQLDIQRKGMWTSATSPTL-----PTPIPTSSVLTAEEK-KARNOALT	4406
Qy	1806	RDREREKSLTSTTVEHAPIWRPGITQSGSSGSSGGGGSSSRP-----ASHSHAH	1858
Db	4407	QAQONQVAVSOVGVGQDIYEFBHDGGEKPKTIISSVAPSAEDQRPRLIITINKTQPSIK	4466
Qy	1859	QHSPISPRTQDALQORPSVLHNTG-----MKGIIITAVEPSKPTVLR	1899
Db	4467	NISEMEQTIQOQOQOQSEVINTDPIGDNSESCNTKSRRLQEKEDRSTVDDIIEDVVR	4526
Qy	1900	STSTSPVRPAATPPPAATHCPLGTLGDGVYPTLMBPVLLPKEAPR-----	1944
Db	4527	NTNT-----PTGTGF--HLPKGAQTP-----PERSGRNAQAKKTDAVQIINA	4566
Qy	1945	VARPERPR-----ADTGHAFKAPPARSGLE-----	1970
Db	4567	VGRPRRSKDRKTIGEOTANLIEVTSASNATVAASHLA--PPEGAGVESHVPQLDAKEVEPV	4625
Qy	1971	----PASSPKSGSRPLPVPVPSGHATIARTPAKNLAPHASDPDPAPPASADPHREKT	2026
Db	4626	SVVTPISITPAPVSAAPVTVFVP-----AMVPVKPTMPQH--PKKXIAAAEIESVQAIN	4678
Qy	2027	QSKPF-----SIQELERLSLGYHSGSYGPEGVEPVSVPSSPLTHDK-----GL	2070
Db	4679	SSIPSGGLPMHQTAAPATQKITGGVADAVSKALVDPVTGVTAGMPQCKEGNLPAAATAA	4738
Qy	2071	PKHLEELD-----KSHLEGEELRPKQP--GPVKLGGEEAAHLPHLRPLPESQPSSSPLL	2120
Db	4739	PANSSNEDGOAAPPLQOHOQOQOHPQOQPPQOQANLQINTTLIPSGLNPINPITGALGSVOL	4798
Qy	2121	QTAPGVKGHORVVTLAQ-HI SEVITYDTR-----HHPQQLSAPL	2159
Db	4799	ETSAALLNKNPVSVLVKGNASQVLIQOQOQPIIVAPAKQPIIILQONPLEPTVLJHHAQHTTVRP	4858
Qy	2160	PAPLYSFGPAGSCPVLDIR-----PPSDBLYLPPPDHGPARGSGPHS-----EGGK	2204

Db	4859	POPL	-----KAYLVNREKNIQOOLTPTKQAVAPPOH-----APHSGHMLLTDTAGNQ	4906
Qy	2205	RSPEPNKTSVLGGEGDIEPVSPPEGTEPCHSR-----SAVYPLLVRDGEQTE	2253	
Db	4907	QLVQPOIIARHLOOQHLOVNVPPPTAHSPHSFRISSQOQQLGFGASISP-----QQQQ	4960	
Qy	2254	PSRMGSKSPGNTSQPPAFFSKLTSNSAMVSKKQEKINKLTHNRNEPYNISQFQTEI	2313	
Db	4961	PQTVVIKQASAAQPOIL--HVVSSKASVVPQOQ--QLPPTSSTGPHLQAKPNYSY	5015	
Qy	2314	F-----NMPAITGTGLM--TYRQAVQEHASTNWGLEAIRKA-----LMGK	2353	
Db	5016	APTVLTTPLFAVQOQQOQHLYKQNNQKGAQIQMPPHGIIMPTHGMLLQOKLPAHLQPQ	5075	
Qy	2354	YDWEESPP-----LSANAFNPLNASASLPAAMPITAAADGRSDHTLTSPGGGGKA	2403	
Db	5076	QHQLNPSPPPGKPNPVLHGLQSGQIMPGSVGSPPPVSAVLKTAQQOVNSVVPVAGIRTA	5135	
Qy	2404	KVSGRPSRKAKSP---APGLASGDRPPSVSSVHSEGD--CNRRT-----PLTN	2447	
Db	5136	IPNISQSQPRVSPVLVPPGISG--VPPFDASLHDLGAVYSGRRTPSPPAHQOASPTP	5193	
Qy	2448	RWEDRSSAGSTFPFYNPLIMLQAG-----VMASPPPGI--PAGSGP---LAGPHHAWD	2499	
Db	5194	NDSTYRGVTAASDFMLYQHILMR--GGDYDDKMGSPPLELRPPGSPPRPTIAVPH---	5247	
Qy	2500	EEPPLLCQVETLSDS	2516	
Db	5248	-----SLOSPQDRTAADS	5260	

RESULT 13

PCLO RAT

ID PCLO RAT STANDARD; PRT; 5085 AA.

AC Q9UKS6; Q9JLT1;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Piccolo protein (Multidomain presynaptic cytomatrix protein).

GN Rattus norvegicus (Rat).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

[1]

RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.

RX MEDLINE=20170257; PubMed=10707984;

RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,

RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;

RT "Piccolo, a presynaptic zinc finger protein structurally related to

RT bassoon".

RL Neuron 25:203-214 (2000).

[2]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

[3]

RP CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;

RP VAL-4668; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND

RP ALA-4694.

RA MEDLINE=21181819; PubMed=11285225;

RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;

RT "An unusual C(2)-domain in the active-zone protein piccolo:

RT implications for Ca(2+) regulation of neurotransmitter release.";

RL EMBO J. 20:1605-1619 (2001).

CC -!- FUNCTION: May act as a scaffolding protein involved in the

CC organization of synaptic active zones and in synaptic vesicle

CC trafficking (By similarity).

CC -!- SUBUNIT: Interacts with Rabac1/Prai, RIMS2 and profilin (By

CC similarity).

CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic

CC junctions.

-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
IsoId=Q9JUKS6-1; Sequence=Displayed;
Name=2;
IsoId=Q9JUKS6-2; Sequence=VSP_003930, VSP_003931;
-!- DOMAIN: C2 domain 1 is involved in binding calcium and phospholipids. Calcium binds with low affinity but with high specificity and induces a large conformational change.
-!- SIMILARITY: Contains 2 C2 domains.
-!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
DR EMBL; AF138789; AAF07822.2; --
DR EMBL; AF227534; AAP63196.1; --
DR HSSP; P04410; 1A25.
DR GO; GO:0045202; C:synaptic junction; IDA.
DR GO; GO:0005509; F:calcium ion binding; IDA.
DR GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.
DR GO; GO:0005522; F:profilin binding; ISS.
DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR008899; Znf_piccolo.
DR Pfam; PF00148; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SF05715; Zf_piccolo; 2.
DR SMART; SM00239; C2; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; 1.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
DR PROSITE; PS50106; PDZ; 1
KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW Repeat; Alternative splicing.
FT DOMAIN 372 491
FT FT
FT ZN_FING 523 547
FT ZN_FING 1010 1033
FT DOMAIN 2351 2362
FT DOMAIN 4442 4536
FT DOMAIN 4653 4752
FT DOMAIN 4968 5059
FT VARSPPLIC 4876 4880
FT FT
FT VARSPPLIC 4881 5085
FT FT
FT MUTAGEN 4668 4668
FT FT
FT MUTAGEN 4674 4674
FT FT
FT MUTAGEN 4688 4688
FT MUTAGEN 4688 4689
FT MUTAGEN 4689 4689
FT MUTAGEN 4690 4691
FT MUTAGEN 4692 4693
FT MUTAGEN 4694 4694
FT SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
12 X 10 AA TANDEM APPROXIMATE REPEATS OF
P-A-K-P-Q-P-Q-P-X.
C4-TYPE (POTENTIAL).
POLY-PRO.
PDZ.
C2 DOMAIN 1.
C2 DOMAIN 2.
TKPTN -> SKRRK (in isoform 2).
/FtId=Vsp_003930.
Missing (in isoform 2).
/FtId=Vsp_003931.
D->A: COMPLETE LOSS OF CALCIUM-BINDING AND CALCIUM-DEPENDENT PHOSPHOLIPID BINDING ACTIVITY.
D->A: COMPLETE LOSS OF CALCIUM-BINDING AND CALCIUM-DEPENDENT PHOSPHOLIPID BINDING ACTIVITY.
V->S: SMALL INCREASE IN AFFINITY FOR CALCIUM.
VM->SS: 10-FOLD INCREASE IN AFFINITY FOR CALCIUM.
M->S: INCREASED AFFINITY FOR CALCIUM.
VV->SS: 10-FOLD INCREASE IN AFFINITY FOR CALCIUM.
QN->AA: MODERATE INCREASE IN AFFINITY FOR CALCIUM.
A->S: NO EFFECT ON CALCIUM-BINDING ACTIVITY.

retina from E11. In the alimentary tract and olfactory epithelium expression was seen from E13. Strong expression present in liver and kidney, from E11 and E13 respectively, and then expression decreased at later stages of development. Moderate expression in lung from E13, while it decreases during postnatal life. Strong expression in thymus from E15 onwards, and in spleen from E17 and during early postnatal life, then, the expression decreases.

-1- DOMAIN: Contains two Leu-Xaa-Xaa-Leu (LXXLL) motifs. Only motif 1 is essential for the association with nuclear receptors.

-1- PTM: Phosphorylated (By similarity).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; AF216186; AAF35860.1; -
DR EMBL; BC031113; AAH31113.1; -
DR EMBL; AF135169; AAF35973.1; -
DR MGD; MGI:192915; Ncoa6.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005667; C:transcription factor complex; IDA.
DR GO; GO:0003682; F:chromatin binding; IDA.
DR GO; GO:0003031; F:retinoid X receptor binding; ISS.
DR GO; GO:0046965; F:retinoid X receptor binding; ISS.
DR GO; GO:0046965; F:thyroid hormone receptor binding; ISS.
DR GO; GO:0003713; F:transcription co-activator activity; ISS.
DR GO; GO:0016563; F:transcriptional activator activity; IDA.
DR GO; GO:0007420; P:brain development; IMP.
DR GO; GO:0001701; P:embryonic development (sensu Mammalia); IMP.
DR GO; GO:0007507; P:heart development; IMP.
DR GO; GO:0030099; P:myeloid blood cell differentiation; ISS.
DR GO; GO:0006367; P:transcription initiation from Pol II promoter; IDA.
DR GO; GO:0006367; P:transcription regulation; Activator; Nuclear protein; Repeat;
KW Transcription regulation; Activator; Nuclear protein; Repeat;
KW Alternative splicing.
FT DOMAIN 1 1060 CREBBP-BINDING REGION (BY SIMILARITY).
FT DOMAIN 1 932 TBP/GTF2A-BINDING REGION (BY SIMILARITY).
FT DOMAIN 1314 NCOA1-BINDING REGION (BY SIMILARITY).
FT DOMAIN 777 931 NCOA1-BINDING REGION (BY SIMILARITY).
FT DOMAIN 1644 2067 EP300/CRSF3-BINDING REGION (BY SIMILARITY).
FT DOMAIN 227 1044 GLN-RICH.
FT DOMAIN 376 381 POLY-PRO.
FT DOMAIN 917 922 POLY-LYS.
FT DOMAIN 1543 1592 SER-RICH.
FT SITE 891 895 LXXLL MOTIF 1.
FT SITE 1495 1499 LXXLL MOTIF 2.
FT VARSPLIC 458 2067 Missing (in isoform 2).
FT MUTAGEN 891 894 /FTId=VSP_003410.
FT CONFLICT 39 39 LVNL->AVNA: ABOLISHES INTERACTION WITH
FT CONFLICT 109 109 NUCLEAR RECEPTORS.
FT CONFLICT 194 194 G -> S (IN REF. 2).
FT CONFLICT 290 290 W -> R (IN REF. 2).
FT CONFLICT 1014 1014 M -> I (IN REF. 2).
FT CONFLICT 1141 1142 Q -> QO (IN REF. 2).
FT CONFLICT 1141 1142 P -> L (IN REF. 3).
FT CONFLICT 1141 1142 SE -> RS (IN REF. 3).
SQ SEQUENCE 2067 AA; 219663 MW; C855F8777167AD48 CRC64;

Query Match 3.38; Score 439.5; DB 1; Length 2067;
Best Local Similarity 19.48; Pred. No. 2.6e-07;
Matches 440; Conservative 295; Mismatches 803; Indels 731; Gaps 111;

591 NSEBAITPQSAELASLMELNESSRWTEEMETAKGLLEHGRNWSAIARVWGSKTVSQCK 650
8 NFEDIYTLSCSTWGDSEVFDSGL--EDDDTKGDSILEDSTIFVAFKGNIDDKD----- 60
651 NFYNYKGRQNLDIRL-----QQHKLKWEKER--NARRKKKAPAAASEAAFPVPV 699
61 ---FWK---LDAILKNVPLNLLHNESSKUKQVQEPWNSVRVTFNIPREAAERLWLAQ 113

700 VEDEMEASGVSGNEEMVEAEALH-----ASGNEVPRGE-----CS 737
114 SNNQRLDGLHSVQ---IEGEGAINLALGONRSQDVRMGVPSVAGNSV-RMEAGFPWAS 169
738 GPATVNNSSDTESTIPSPHTEAAKOTGQNGPKPPATLGDGPGPPPTPPRTSRAPTEPT 797
170 GFLGIRMTS-----PAAVMTPOGNNSSNM---AFGNPELOPR-----TPR 209
798 PASEA-----TGAPTPPPAPPSAPP-----PVVPKBEKEET 831
210 PASQSDAMDPLLSGLHIQQSHPSGSLPPAHHSMPQPVVNRQMNPAFPQLQQQQQQQQ 269
832 AAAPVVEEGEEKPPAAAEALAVDTGKAEPEVKSECTEAEER-----PAKGKDAEAE 884
270 QQQQQQQQQQQQQQQQQQQQQQ---RPLQH---QQOQIGIRPQFTAPTQVPPVPGWN 323
885 ATAEGALKAEKKEGGSGRATTAKS--SGAPQSDSDSATCSADEVDEAGDKNRLSLRPS 943
324 QLPFGALQPPPAQGSIGTITNQGWKAPLPSPMAQ-----LQARPS 366
944 LITPTGDPANASPOKFLDLKQKORAAAIPIQVTKVHEPPREDAAATKPAAPPAPPQ 1003
367 LAT-----VQT-----PSHPPPPYF 382
1004 NLQPESDAPQ--PGSSPRGKSRSPAPPADKEAFAPAAQKLPDPPCWTSGLPFPVPPREV 1062
383 GSQQAQAHTNFPQMSNPGQFTAP-----CMKGLOGGP---SRVFTPLQQPHL 427
1063 IKASPHADPPSPSY-APGCHPLPLGLHDTARVPLPRPTISNP-----PPLISSAKHP 1115
428 TNKSP-ASSPSPSQGSPASSPT---VNOQOQMGPRPQ--NNPLSQGQQQVSSPGRNP 482
1116 SVLEROIGALSQMSVOLHVPSSEHAKAPVGPVTGMLPLPMDPKKLAPSGVKQEQLSPR 1175
483 MV---QCGNVPPFMVWQQPPNQ-----GPOSLHPLGLGMPKRLPP--GFSAGANPN 531
1176 GQAGPPESLVPTAQEASVLRGTALGVSVP--GGSITGIGISTRVPSDSAITYRGSITHTP 1234
532 FMQGVPTTAAATPGNSGALQLOANQNVHAGQGGAGPPQNM-----QVSHGPP 581
1235 ADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVIEGKKGHLYSVEGMS--VTQCS--K 1291
582 NNM-----QPSLM-----GIHGINNQAGSGVGPVQTLGN 612
1292 EDGRSSSGPPHETAAPKRTYDMMEGRVGRASIASIEGLMGRAPIPERHSPHLKQHHI 1351
613 MQGPPQQGPP-----SOLMGMEQQIVP-----SOGQMAQQ--- 642
1352 RGSITQIGIPRSYVEAOEDYLARREKLLA---KREGTPPPP---PPSRDLTE-----AYK 1398
643 QGTLNQNPMILSRAQ---LMPQGMVNNAQNQLGSPQBMTPPKQMLPQCGPQMMAHP 699
1399 TOALGP-----LKLKPAHEGLVATVKEAGRSIHIIPREE--LRHTPELPLAPRLKEGSI 1451
700 NQMMPGOVQLQONPMIEQIMTNMQGNKAQNFNSQNSVMMPGPAQIMRGPTTNMOGNN 759
1452 TQGTPLKYDTGASTTGSKHVDVRSIGSPGRTPTTPVHPLDVMADARALERACYEESLSKR 1511
760 VQFTQMSQWMLPQCGPVNNSPSQVMGICQGVLRPPGSPHMAQQ-----TD 807
1512 PCTASSSGSITARGAPVI-----VPELKGPRQ-----SPLTYEDHG-----APFAGHLPR 1556
808 PVTANNDVNLQMMPPDVSMQASNVPPHVSQMGNASGSHFSGHGVSNAPFGG-APN 866
1557 GS-----PVTMREP--TPRLQEGSLSSKASQDRKLTST-----PREI 1592
867 GSQMSCGQNPFPVNVKDVTLTSLPLVNLQSDISAGHFGVNNKQNTNANKPKKKKPPRK 926
1593 AKSPHS--TVPEHHHPISPY-EHLLRCVGVDLVRSIPLAFDPTSIPIRGIPLDAAAY 1649
927 KKNCHQDLNTPNRPRTGLEEDVQQLPGEQGINLDTTG-PKLPDFSNRPVGYPTQVPEQR 985

Qy	1650	YLP-----RHLPNPTYPHLYP-PYLIR-----GYDPTAALENROTIINDYTTSQQ	1691
Db	986	PLPQPMPQLMHVAPPQPPOQOQPQLPQOOQPPFSPQSQOQOQOQMM-MLMMAQ	1044
Qy	1695	MHNTATAMAQRADWLRG-LSPRESSLALNYAAGPRGIIDLQSVPHLPVLVPPTP----	1748
Db	1045	DPKSIRLPVSQNVHPGRLPNDSORMPVOQSGNVPMVGLOG-----PASVPSPDKORM	1100
Qy	1749	----GTPATAMDRLAYILTPAPOFFSRHSPLS-----PGGPTHLTKP	1788
Db	1101	PMSVNTWGSNRKWVQENPQ-----NSSSSPLCEMSSLPEASGVSPSVAGGNPNMPSH	1156
Qy	1789	TTTSSRRERDRDRDREREKSLITSTTVEHPIWRPGTEQSSGSSGGGGSS	1848
Db	1157	LVSQNO-----LMMTGP-----KPGSPLSATOGATQOPPVN	1190
Qy	1849	SRPASHGAHQHSP-ISPRTODALQORPSPVLNHTMGKIITAVEPSKPTVLRJSTSTS-SP	1906
Db	1191	SLPSSHGH---HFVNVAAPTQS---RKPTPNRRSPRYYPQTNNRPSTEPSEISLSP	1244
Qy	1907	VR-----PAATPPPPTHCEL-----GGTLDGVTPLMEPV-----LLPKCAPRVARP	1948
Db	1245	ERLNASTAGLFPPOINTELPPRPNLNGFDQGLNLPTLLKAIGQAPSNLITITPPNPFAAP	1304
Qy	1949	ERPRADT-----GHAFIAKPPARSGLEPASS-----PSKGSS-----	1980
Db	1305	QAHKLDVVNSGKOSQNPGITKRASPNSRRSSFGSSRKTTTPSPGRQNSKAPKLTLASQT	1364
Qy	1981	-----PRLPVPVSGHATIARTPAKNLAPHHA--SPDPAPPASA-----	2018
Db	1365	STTMLQNMELPRNV---LVGPTPLANPLPGSFNNNTGLNPQNPTVPVPMANGTVLEDNKE	1421
Qy	2019	-----SD-----PHREKTQSKPFSIOEELERSLGXGSSVSPGGEVPSVPSSPSL	2064
Db	1422	SVINPDSDCONAGRKEQWNTELKVVTQEAKNAV-----PEQSKKDGOPLDNKLPVS	1477
Qy	2065	THDKGL-----PKHLEELDCKSHLEGERLKPOG-----PVKLGGE-----AAHLPH	2105
Db	1478	EENKNLMSPAWREAPTSLSQLLDNSGAPNVTIKPPLGLTDLEVTPPVVSGEDLRKASVIPT	1537
Qy	2106	LRPLPESPQSSPLLQTAPGVKHQHVVTIAQ-HISEVITQDYTRHHPOQLSAPLPAPLY	2164
Db	1538	LQPPSPKEPSTS-----LSSP-HGSEPCTLIARELSEVSNN-----AAPSIPPPVM	1582
Qy	2165	SFGASCPLDLRRPPDLPLPPDHCAPARGSPHSEGGKRSPEPNKTSVLGGEGDIGEP	2224
Db	1583	SRPVSS-----SSIETPLP-----PNOITVFVTS-----NP	1608
Qy	2225	VSPPEGWTE---PGHSRAVPLYVRDGEQTEPSRMGSK-----SPGNTSOPPAFP SKL	2275
Db	1609	ITTSNSTSAALPHLOQALMSTV---TWPN-VGNKMVVBESQSAQAQSNARPQFITPV	1662
Qy	2276	TENSAMVKSKQOEINKKLATHNRNEPEYNISOPGTEIFNMPLTGTLGLMYTYSQAOE-	2334
Db	1663	FINSSTIIQWKG-----SQPST-IPATPLTTSGLMPPSVAVVGPL	1703
Qy	2335	HASTNMGLEAIIRKALMGKYDOWEESPPLSANAFNPLNASA-SLPAMPITAADGRSDHT	2393
Db	1704	HIPQNI-----KF-----SSAPVTPNV---PSSSPAPNIQTGRPLVLSRATPVQ	1745
Qy	2394	LTPSGGGGKAKVSGRPSRKA KSPAPGLAS-----CDRPPS-----VSSV---HSEG	2437
Db	1746	LPSPPTCTSPVWAPNFSPVOVKELNDPEASPTQNTSADQSTLPPSPQTTVVVSLLTNSPG	1805
Qy	2438	-DCNRRRTPLTNVWEDRPPSAG-----	2458
Db	1806	SSANRRSPVSSKKGKVDKIGILLTKACKVYTGSLKEGEQYGADGETEGPGLBITTP	1865
Qy	2459	-----STPFPPYNPLIMRLQAGMA-----SPPPPGLPAGSGP	2490
Db	1866	GLMGTEQCSTELDSKTPSPAPTLTKKTSZFMAPSSSTGTGPILPGGALP	1914

RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ichii S., Kawai Y., Saito K.,
RA Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 3879-4911 FROM N.A.
RC Tissue=Testis;
RA Duesterhoef A., Lauber J., Mewes H.-W., Weil B., Wiemann S.,
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX (ISOFORM 2).
RC Tissue=Cervical carcinoma;
RX MEDLINE=22371496; PubMed=12482968;
RA Goo Y.-H., Sohn Y.-C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
RA Azorsa D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C.,
RA Lee J.W.,
RT "Activating signal cointegrator 2 belongs to a novel steady-state
RT complex that contains a subset of trithorax group proteins.";
RL Mol. Cell. Biol. 23:140-149(2003).
CC -!- FUNCTION: Belongs to the ASC-2/NCOA6 complex (ASCOM), a
CC coactivator complex of nuclear receptors, involved in
CC transcriptional coactivation. MLL3 may be a catalytic subunit of
CC this complex, which weakly methylates Lys-4 of histone H3. This is
CC a specific tag for epigenetic transcriptional activation. May be
CC involved in leukemogenesis and developmental disorder.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =
CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
CC -!- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/
CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins
CC MLL2 and MLL3, and ASH2/ASCL2. Interacts with histone H3.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8NEZ4-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8NEZ4-2; Sequence=VSP 008561, VSP 008562;
CC -!- TISSUE SPECIFICITY: Highly expressed in testis and ovary, followed
CC by brain and liver. Also expressed in placenta, peripheral
CC blood, fetal thymus, heart, lung and kidney. Within brain,
CC expression was highest in hippocampus, caudate nucleus, and
CC substantia nigra. Not detected in skeletal muscle and fetal liver.
CC -!- DOMAIN: The SET domain interacts with histone H3 but not H2A, H2B
CC and H4, and may have a H3 lysine specific methylation activity.
CC -!- MISCELLANEOUS: Found in a critical region of chromosome 7, which
CC is commonly deleted in malignant myeloid disorders. Partial
CC duplication of the MLL3 gene are found in the juxtacentromeric
CC region of chromosomes 1, 2, 13 and 21. Juxtacentromeric
CC reshuffling of the MLL3 gene has generated the BAGE genes.
CC -!- SIMILARITY: Belongs to the TRX/MLL family.
CC -!- SIMILARITY: Contains 1 DHHC-type zinc finger.
CC -!- SIMILARITY: Contains 6 PHD-type zinc fingers.
CC -!- SIMILARITY: Contains 1 post-SET domain.
CC -!- SIMILARITY: Contains 1 SET-type zinc finger.
CC -!- SIMILARITY: Contains 1 SET domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AY024361; AAK00583.1; -;
CC EMBL; AF264750; AAF74766.2; -;
CC EMBL; AC006017; AAD45822.1; -;
CC EMBL; AC104692; -; NOT ANNOTATED_CDS.
CC EMBL; AC005631; -; NOT ANNOTATED_CDS.

DR EMBL; AB040939; BAA96030.2; -;
DR EMBL; AK022687; BAB14179.1; -;
DR EMBL; AK075113; BAC11409.1; -;
DR EMBL; AL833924; CAD38780.1; -;
DR Genew; HGNC:13726; MLL3.
DR MIM; 606833; -;
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR003889; FYZich_C.
DR InterPro; IPR003888; FYZich_N.
DR InterPro; IPR000910; HMG 12 box.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR001214; SET.
DR InterPro; IPR001594; Znf DHHC.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_Ring.
DR Pfam; PF00505; HMG_box; 1.
DR Pfam; PF00628; PHD; 6.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00542; FYRN; 1.
DR SMART; SM00541; FYRN; 1.
DR SMART; SM00398; HMG; 1.
DR SMART; SM00249; PHD; 8.
DR SMART; SM00508; PostSET; 1.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS00354; HMG1_Y; 1.
DR PROSITE; PS00868; POST_SET; 1.
DR PROSITE; PS50280; SET; 1.
DR PROSITE; PS50216; ZF DHHC; 1.
DR PROSITE; PS01359; ZF_PHD_1; 5.
DR PROSITE; PS00016; ZF_PHD_2; 6.
DR PROSITE; PS50089; ZF_RING_2; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
DR Transferrase; Methyltransferase; Chromatin regulator; Activator;
DR DNA-binding; Nuclear protein; Transcription regulation; Coiled coil;
DR Zinc-finger; Repeat; Alternative splicing; Polymorphism.
FT ZN_FING 341 391 PHD-TYPE 1.
FT ZN_FING 344 389 RING-TYPE.
FT ZN_FING 388 438 PHD-TYPE 2.
FT ZN_FING 436 489 DHHC-TYPE.
FT ZN_FING 464 520 PHD-TYPE 3.
FT ZN_FING 957 1010 PHD-TYPE 4.
FT ZN_FING 1007 1057 PHD-TYPE 5.
FT ZN_FING 1084 1139 PHD-TYPE 6.
FT DOMAIN 4770 4891 POST-SET.
FT DOMAIN 4895 4911 COILED COIL (POTENTIAL).
FT DOMAIN 92 112 COILED COIL (POTENTIAL).
FT DOMAIN 644 672 COILED COIL (POTENTIAL).
FT DOMAIN 1338 1366 COILED COIL (POTENTIAL).
FT DOMAIN 1754 1787 COILED COIL (POTENTIAL).
FT DOMAIN 3054 3081 COILED COIL (POTENTIAL).
FT DOMAIN 3173 3272 COILED COIL (POTENTIAL).
FT DOMAIN 3391 3433 COILED COIL (POTENTIAL).
FT DNA_BIND 34 46 A.T HOOK (BY SIMILARITY).
FT DOMAIN 1719 1796 GLN-RICH.
FT DOMAIN 1834 2281 PRO-RICH.
FT DOMAIN 2412 2630 PRO-RICH.
FT DOMAIN 2690 2786 ASP-RICH.

Query Match 3.3%; Score 439; DB 1; Length 4911;
Best Local Similarity 18.3%; Pred. No. 6e-07;
Matches 556; Conservative 313; Mismatches 1010; Indels 1152; Gaps 142;

QY 34 HTDCLVLEVQHSDRYASHLSFGSIQIP---QRRRPSLLSEFPQFNERSQELHLRPESHHS 90
Db 1458 HSDIGFVTD-----DPSSLPQPNVNSRP--LSEQLDGLSPDLKXVTDGA 1504
QY 91 YL-----PELGKSEMEFIESKRPLELLPDLRPSPLLATGQPGASEDTKDRSLTGK 144
Db 1505 ILGLKYKIPELGKGVDEL-----FTAVLSPANT----- 1533
QY 145 LEFVSPFPPTDTPLELVPPRLSKKELIQNMDRVDRITWVQQIISKLKKKQOOLEBEA 204
Db 1534 -QFTPLPQPP-----PPTQLLP-----IHNQDAFSR-WPLMNGLI-----G 1567

RA	3292	PGATPTMSQTFMVPQQLHQOQHTTVISGHTSPVRMPSLPQWQNSAPAHLPPLNPRI	3351	RA	Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RT				RT	Wohlman P.;
RT	2013	-----APPASADPHREKTKQKPF-----	2032	RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT				RT	elegans.";
RL	3352	QPPIAQLPIKCTPAGCTVSNANFQSGPPPPVDFDNNPFSFQERKERLREQOERQ	3411	RL	Nature 368:32-38(1994).
RN				RN	(2)
RP	2033	-----IQELERLSLGHGSSVSP-----	2061	RP	REVISIONS, AND ALTERNATIVE SPLICING.
RA				RA	Waterston R.;
RL	3412	RIQLMQEVDQRALQORMEMEGHGMVGEISSSRTSVSQIPFYSSDLPCDFMPLGLQ	3471	RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC				CC	- - SUBCELLULAR LOCATION: Nuclear (Probable).
CC				CC	- - ALTERNATIVE PRODUCTS:
CC				CC	Event=Alternative splicing; Named isoforms=2;
CC				CC	Name=b;
CC				CC	isoId=p34333-1; Sequence=Displayed;
CC				CC	isoId=p34333-2; Sequence=VSP 003306, VSP 003307;
CC				CC	Note=No experimental confirmation available;
CC				CC	- - SIMILARITY: Contains 1 Myb-like domain.
CC				CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC				CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC				CC	the European Bioinformatics Institute. There are no restrictions on its
CC				CC	use by non-profit institutions as long as its content is in no way
CC				CC	modified and this statement is not removed. Usage by and for commercial
CC				CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC				CC	or send an email to license@isb-sib.ch).
CC				CC	-----
DR		EMBL; L15188; AAA27949.3; -		DR	EMBL; L15188; AAA27950.1; -
DR		PIR; S44758; S44758.		DR	PIR; S44758; S44758.
DR		WormPep; C14B9.6a; CE00077.		DR	WormPep; C14B9.6b; CE29577.
DR		InterPro; IPRO01005; Myb_DNA_binding.		DR	Pfam; PF00249; myb_DNA-binding; 2.
DR		SMART; SM00717; SANT; 2.		DR	SMART; PS00337; MYB 1; FALSE_NEG.
DR		PROSITE; PS00334; MYB 2; FALSE_NEG.		DR	PROSITE; PS00334; MYB 2; FALSE_NEG.
DR		PROSITE; PS00590; MYB 3; 1.		DR	PROSITE; PS00590; MYB 3; 1.
KW		Hypothetical protein; DNA-binding; Nuclear protein;		KW	Hypothetical protein; DNA-binding; Nuclear protein;
KW		Alternative splicing.		KW	Alternative splicing.
FT		DNA_BIND 436 486		FT	DNA_BIND 436 486
FT		VARSPLIC 1019 1780		FT	VARSPLIC 1019 1780
FT		Missing (in isoform a).		FT	Missing (in isoform a).
FT		/FTId=VSP_003307.		FT	/FTId=VSP_003307.
FT		SEQUENCE 1780 AA; 198569 MW; 0364817C41A8AD3F CRC64;		FT	SEQUENCE 1780 AA; 198569 MW; 0364817C41A8AD3F CRC64;
Qy		Query Match		Qy	Query Match
Qy		Best Local Similarity		Qy	Best Local Similarity
Qy		Matches 423; Conservative 281; Mismatches 737; Indels 779; Gaps 104;		Qy	Matches 423; Conservative 281; Mismatches 737; Indels 779; Gaps 104;
Qy	270	IKINQAMRKLLT-----LYFKRNHARKQWKQKFCQRYDQLEALEKKVERIENN	319	Qy	66
Qy	66	LKVNAGDLLLIFCSFPCLFDVLVFTFKELARE-----YDEKFAFQDDI-QKWEET	116	Qy	320
Qy	320	PRRRAKESKREYVEYKOPFEIRKQELQRMQSRVQGRGSLMSAARSEHEVSEIDGL	379	Qy	117
Qy	117	SERKEYAEFHRVQAESEFPPELKRERERW-----AAAEIRGEDEKSM	161	Qy	380
Qy	380	SEQUENLEKQMR-QLAVIPPMLYDADQOIRKFINNGLMADPMKYKDVNM---NSEQ	435	Qy	162
Qy	162	LAKEHADKKIRLGVAKIPRLITSESKQVDEFEVERPGSILKDMKKKHEHROSULDRLEWSP	221	Qy	436
Qy	436	EKETREKFMQHPKNGFIAGFLERKTVARCVLYITLKQKNYKSLVRRSYRRGKSQQ	495	Qy	222
Qy	222	ERSLFKSROADHVKIFHGLTFFVDKTDASDLVLFYNNKKTEDYKDKPKKRVTKYKVG	281	Qy	496
Qy	496	QQQQQQQQQQQQQQQQMPRSSQEEK-----DEKEKEAEK-----EE	533	Qy	282
Qy	282	APPSVEELAYFRMPPPLDFSSPPKNSLMCYFCCTRWNGIDLNGTTPMPKEAYEIFAICPDE	341	Qy	

RP TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.
RC TISSUE=Brain;
RX MEDLINE=9439764; PubMed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Kilimann M.W.;
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin";
RL J. Cell Biol. 147:151-162(1999).
RN [2]
RP REVISIONS.
RA Kilimann M.W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 4502-4682 FROM N.A.
RC TISSUE=Brain;
RA Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP INTERACTION WITH RIMS2.
RX MEDLINE=22384373; PubMed=12401793;
RA Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,
RA Sasaki T., Tajima N., Iwanaga T., Seino S.;
RT "Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of
RT CAMP-GEFII.Rim2.Piccolo complex in cAMP-dependent exocytosis";
RT J. Biol. Chem. 277:50497-50502(2002).
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking.
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin.
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9QYX7-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9QYX7-2; Sequence=VSP_003929; VSP_003929;
CC -!- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in
CC stomach. Not detected in other tissues analyzed including adrenal
CC gland, testis and pancreas.
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y19185; CAB60731.2; --
CC EMBL; Y19186; CAB60732.2; --
CC EMBL; AF181269; AAD55786.2; --
CC HSP; P04410; 1A25.
CC MGD; MGI:1349390; Pclo.
CC GO; GO:0045202; C:synaptic junction; IDA.
CC GO; GO:0005509; F:calcium ion binding; ISS.
CC GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
CC GO; GO:0005522; F:profilin binding; IDA.
CC GO; GO:0019933; P:CAMP-mediated signaling; IDA.
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.
CC GO; GO:0030073; P:insulin secretion; IDA.
CC GO; GO:0017157; P:regulation of exocytosis; IDA.
CC GO; GO:0016080; P:synaptic vesicle targeting; NAS.
CC InterPro; IPR000008; C2.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR008899; Znf_piccolo.
CC Pfam; PF00168; C2; 2.

DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF05715; Zf_piccolo; 2.
DR SMART; SM00239; C2; 1.
DR SMART; SM00238; PDZ; 1.
DR PROSITE; PS00499; C2 DOMAIN 1; 1.
DR PROSITE; PS00004; C2 DOMAIN 2; 2.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW Repeat; Alternative splicing.
FT DOMAIN 371 470
FT P-A-K-P-Q-P-Q-P-X.
FT ZN_FING 502 526
FT ZN_FING 967 990
FT C4-TYPE (POTENTIAL).
FT C4-TYPE (POTENTIAL).
FT POLY-PRO.
FT PDZ.
FT C2 DOMAIN 1.
FT C2 DOMAIN 2.
FT C2 DOMAIN 2.
FT TRPTN -> SKRRK (in isoform 2).
FT /FTID=VSP_003928.
FT Missing (in isoform 2).
FT /FTID=VSP_003929.
FT VARSPLIC 4834 5038
FT SEQUENCE 5038 AA; 547600 MM; DADA460CF3B40888 CRC64;
SQ
Query Match 3.3%; Score 434.5; DB 1; Length 5038;
Best Local Similarity 19.8%; Pred. No. 8.6e-07;
Matches 391; Conservative 216; Mismatches 701; Indels 663; Gaps 91;
688 AAASEEAPPPVVEDEMEASGVSGNEEMVEAEALHASGNEVPRGECSPATVNNSSD 747
23 AGSGGALHPGIPAGMEADLSQLS---EEERQIAAVMSRAQGLPKG--SVPAAMAAES-- 75
748 TESIPSPHTEAAKDTGQNGPKPATILGADGPPGPPPTPRR-TSRAPIEPTPASEATGAP 806
76 ----PSMHRKQELDSQ-APQQP-----GKPPDPGRPFQHGLSKSRITDTFRSE---OK 121
807 TPPAPPSPAPPPVVPVVEKEEETAAAPVEBG--EEOKPPAAEELAV-----DTGKA 858
122 LPCRSPSTILSKSKSRITDPKEEYKSMMP---GPFSEVNPLSAVSVVKNFNPDLISD 178
859 EEPVKSCTEAELEGPAKGDAEAAEATAEALKAEE-----KKEGSGRATTAKSS 909
179 SEAQVEETTK-----KQVAKQDQKSGEIGTKPSLQSPKLI1PKQQPGKEV----- 226
910 GAPDSSSATCSADEVDEAGGDKNLLSP---RSLLTPTGDPANASFP--OKPLDLK 964
227 -IPODIEFSKVS-----QAEKTKPQAPGTAKPSQSPAQTPAQQAQKFAVQAQGPAP 278
965 -QLKQRAAATPIQVTKVHEPPREDAAPTKAPPAPPPQNLOPESDAPQPGSS----- 1018
279 ATVOQPGPKSPAQ-----PAGTKGSPAQPPVTAKPPAQAGLEKTSLOQPGPKSLAQ 332
1019 -----PRGKSRSPAPPADKEAFAAE-----AQKLPG--DPPCWTSLGPPVP----- 1058
333 PGQGVPPGPAKSPAQPGTAKLPAQPGPGTAKVFGTKTTPAQLSG-PKPTPAQPGP 391
1059 ----PREVIKASHPAPPSAFSYP-----PGHPLPLGLHDTARVPLPRE--PTIS 1103
392 TKPSQGPPIPAKPOQPPVATKPOQPPAPAKPQFQHTPA----KQPQOQTPAKPOQ 447
1104 NPPPLISSAKHPSV-----LERQIGAIISQMSVQ-LHVPYSEHAKAPGVPTMGLPLPMDP 1158
448 QFTPAKPOQHPGLKPKSAQPSKISQTVTGRPLQAPPTSAQAQ-----AQGL-----S 498
1159 KKLAPFSGVKQEQLSPRQAGPPESLGVPTAQEASVLGALSGVPGSGITK----- 1210
499 KTCICPLCNTTELLH-----TPEKANFNTCTEQSTVCSLCGNFNPPLHTEIKEWLCLN 552
1211 -----GIPSTRVPSDSAITYRGSIHTGTPADVLYKGTITRIIGSDSSRLDRG 1258
553 CQMQALGGELAAIFSSPQTPKAASVQPATASKSPV-----PSQQAAP 596
1259 REDSLPKGHVIYEGKKHVLSEYEGMSTVQCSKEDGRSSSGPPHETAAPKRTYDMWEGRV 1318
::: |||

Db 597 KKE-LP-----SKQD-----SPK----- 608
Qy 1319 GRAISSASIEGLMGRAIPERHPHLLKEQHHIRGSITQGIPIRSYVEAQEDYLREAKLL 1378
Db 609 -----APESKKPPLPVKQPTLHGPTATAPQPPV----- 637
Qy 1379 KREGTPPPPSRDLTEAYKTOALGPL-CLKPAHEGLVATVKEAGRSIHEIPREELRHTP 1437
Db 638 -AEALPKPAPPKP-SAALPEQAKAPVADVEPKQPTTETLTDSPSSAAATSKPAI----- 691
Qy 1438 ELPLAPRLKGSITQGTPLKYDTGASTTGSKKHVDVRSGLIGPGRTFPPVHPLDVMADAR 1497
Db 692 ---LSSVQAQAQVTTAPLTKDTSKTS-----OSFPPTGDTTTPLDISK 732
Qy 1498 ALERACYBESLKSREGTSSSGSSTARGAPVIVPELKGPRQSPLTYEDHGAPFAGHLPRG 1557
Db 733 AMPRPASDSKIVSHEGPTSESKD-----PVQKKEPKKAQTKVTPKPTDKP-----VPKG 782
Qy 1558 SPV-----TWREPTPRLOEGLSSSKASQDKLTSTPRETAKSPHSTVPEHHPIPSY 1611
Db 783 SPTSPGTRPTTGOATPOSQO---PPKPPQESRRFSLNLGGIADAPKS-----QPTTPQ 832
Qy 1612 EHLRGVSGVDLYRSHIPLADPTSI PRGIPLDAAAAYLPHLAPNTYPLHYPPYLIR 1671
Db 833 ETVTGKLPF-----FGASIFSQASNLISAGQOAP-HPQTGPAAPSKQAP----- 876
Qy 1672 GYPDTAALLENQTIINDYITSQOMHHNTATAMAQADMLRGLSPRESSLALNYAAGPRGI 1731
Db 877 -----PSSQTLA---AQGP--- 887
Qy 1732 IDLSQVPHLVLPPTGPTATAMDRLAYLTPAQPFSSRRSSPSLPGGTHLTKPTTT 1791
Db 888 -----PKSTG-----QHPSAKATKAVKTKG 910
Qy 1792 SSSERDRDRDRDRREKSILSTTTVEHAPIWRGTQSSGSSSSSGGGSSSRP 1851
Db 911 PAENLEAKPAQAPTVKKAED-----KKHPP-----GKVS KP 943
Qy 1852 ASHSHAHQHSIPRTQDALQORSVLHNTGMKGIITAVEPSKPTVLKSTSTSPVRPAA 1911
Db 944 -----PTEPE-KAVLAQKPDK-----TTXPKPACPLCRTELNVGQDPP-- 982
Qy 1912 TFPATHCP-IGGTLGDVYPT-----LMPEVLLPKAPRVARPERPRADTGHAFLAKPPA 1965
Db 983 NFNTCTECKNOVCNLGNFPTPHLTEIQEWCLNCQTORAISGQLGDM-----KMPPA 1036
Qy 1966 RSGLEPASPCKG-SERPPLVPPVSGHA-----TIATPAKNLAPHASPD-PPAPPAS 2017
Db 1037 SSG--PKASPVPAFAEPFPQKPTTAAHAHAKKKKTEVKAETEKIPEKETESIEKTPPAV 1094
Qy 2018 ASDPHREKTQSKPFSIQELELSLGYHGSSYSPEGVPEVPVSPVPSLTHDKGLPKHLEL 2077
Db 1095 ATD-----QKLE-----ESEVTKSLVSVLPEKPS-EEKALPADKKE- 1131
Qy 2078 DKSHLELGRPKQPGVKVKGGEAAHLPHLRPLPESQ-----PSSPLLQATPCVKGHQ 2130
Db 1132 -----KXPP-----AAEAPPLEBKPIPDQKLPDAKPSASGEKRDLLKAHV 1176
Qy 2131 RVV-----TLAQHSEVITQDTRHHHQOLSAFLPAPLYFPGASCPLV---DLRRPPSD 2182
Db 1177 QIPEEGPIGVKSLACGEQOPDTR--PEDL-----PGATPQTLPKDRQKESRD 1223
Qy 2183 LYLPPPHDGA-PARGSPHSSEGGKSPENKTSVLGGGEDGIEPVSPPEGMTEPGHRSASV 2241
Db 1224 VTQPAEGTAKEGGEFSKDRTEKEDSKDTS-----SSQPKS-POGLSDTGYSSDGI 1276
Qy 2242 -----YPLLVRDGE----- 2250
Db 1277 SGSLGEIPLSPIDEKDLLGLKKDPSQESSPSSPSDLAKLESTVLISLEAQASTLVGE 1336
Qy 2251 -----QTPSRMGSKSPGN--TSQPAFPKSLTENSAMVSKKEINKLNT----- 2296
Db 1337 KAEEKTOPKVPQPODQOQKOTOTPSETRDISIEEIKESQEKVTSKKDSAQGFPSRK 1396

Qy 2297 HNRNEPE-----YNISQPGTEIFNMPAI-----TGTGLMYRSQAVQEHASTNMG 2342
Db 1397 EHKENPELVDDLSPPRASYSVDSESSSESPVARRKRTTSIGSSSSSEYKQEDSQSGE 1456
Qy 2343 -EAIIRKALMKYQOWEESPPLSANAFNPLNASASLPAAPIAAGRSDHDLTSPGGGG 2401
Db 1457 DEDFIRKQII-----EMSADE-----DASGS-----EDEFIRSQLKEIGV 1493
Qy 2402 KAKVSGRPSSRKAKSPAGLASGRPPSVSVHSEGDGCDNRRTPLTNRWED 2452
Db 1494 TESQKRETKGKSPKPA---GKHRLTRKSTSFDDAGRR-----HSWHD 1536
RESULT 19
SHK1 RAT
ID _SHK1 RAT STANDARD; PRT; 2167 AA.
AC Q9WV48; Q9QZ28; Q9WU13; Q9WUE8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP
interacting protein) (SPANK-1) (Synamon) (Somatostatin receptor
interacting protein) (SSTR interacting protein) (SSTRIP).
GN SHANK1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND
DLG4.
RP TISSUE=Brain;
RC MEDLINE=99419021; PubMed=10488079;
RX Yao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
"Synamon, a novel neuronal protein interacting with synapse-associated
protein 90/postsynaptic density-95-associated protein.";
J. Biol. Chem. 274:27463-27466(1999).
[2]
RP SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=99360650; PubMed=10433268;
RA Naibitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,
Weinberg R.J., Worley P.F., Sheng M.;
"Shank, a novel family of postsynaptic density proteins that binds to
the NMDA receptor/PSD-95/GKAP complex and cortactin.";
Neuron 23:569-582(1999).
[3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20549637; PubMed=10958799;
RA Tobaben S., Suedhof T.C., Stahl B.;
"The G protein-coupled receptor C1i interacts directly with proteins
of the shank family.";
J. Biol. Chem. 275:36204-36210(2000).
[4]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND
DEVELOPMENTAL STAGE.
RC TISSUE=Brain;
RX MEDLINE=99436166; PubMed=10506216;
RA Lim S., Naibitt S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;
"Characterization of the Shank family of synaptic proteins. Multiple
genes, alternative splicing, and differential expression in brain and
development.";
J. Biol. Chem. 274:29510-29518(1999).
[5]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
"Somatostatin receptor interacting protein defines a novel family of
multidomain proteins present in human and rodent brain.";
J. Biol. Chem. 274:32997-33001(1999).

[6] INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
MEDLINE=99360651; PubMed=10433289;
Tu J.C., Xiao B., Naibitt S., Yuan J.P., Petralia R.S., Brakeman P.,
Dean A., Akalu V.K., Lanthan A.A., Sheng M., Worley P.F.;
"Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
postsynaptic density proteins.";
Neuron 23:583-592(1999).
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
[17]
[18]
[19]
[20]
[21]
[22]
[23]
[24]
[25]
[26]
[27]
[28]
[29]
[30]
[31]
[32]
[33]
[34]
[35]
[36]
[37]
[38]
[39]
[40]
[41]
[42]
[43]
[44]
[45]
[46]
[47]
[48]
[49]
[50]
[51]
[52]
[53]
[54]
[55]
[56]
[57]
[58]
[59]
[60]
[61]
[62]
[63]
[64]
[65]
[66]
[67]
[68]
[69]
[70]
[71]
[72]
[73]
[74]
[75]
[76]
[77]
[78]
[79]
[80]
[81]
[82]
[83]
[84]
[85]
[86]
[87]
[88]
[89]
[90]
[91]
[92]
[93]
[94]
[95]
[96]
[97]
[98]
[99]
[100]
[101]
[102]
[103]
[104]
[105]
[106]
[107]
[108]
[109]
[110]
[111]
[112]
[113]
[114]
[115]
[116]
[117]
[118]
[119]
[120]
[121]
[122]
[123]
[124]
[125]
[126]
[127]
[128]
[129]
[130]
[131]
[132]
[133]
[134]
[135]
[136]
[137]
[138]
[139]
[140]
[141]
[142]
[143]
[144]
[145]
[146]
[147]
[148]
[149]
[150]
[151]
[152]
[153]
[154]
[155]
[156]
[157]
[158]
[159]
[160]
[161]
[162]
[163]
[164]
[165]
[166]
[167]
[168]
[169]
[170]
[171]
[172]
[173]
[174]
[175]
[176]
[177]
[178]
[179]
[180]
[181]
[182]
[183]
[184]
[185]
[186]
[187]
[188]
[189]
[190]
[191]
[192]
[193]
[194]
[195]
[196]
[197]
[198]
[199]
[200]
[201]
[202]
[203]
[204]
[205]
[206]
[207]
[208]
[209]
[210]
[211]
[212]
[213]
[214]
[215]
[216]
[217]
[218]
[219]
[220]
[221]
[222]
[223]
[224]
[225]
[226]
[227]
[228]
[229]
[230]
[231]
[232]
[233]
[234]
[235]
[236]
[237]
[238]
[239]
[240]
[241]
[242]
[243]
[244]
[245]
[246]
[247]
[248]
[249]
[250]
[251]
[252]
[253]
[254]
[255]
[256]
[257]
[258]
[259]
[260]
[261]
[262]
[263]
[264]
[265]
[266]
[267]
[268]
[269]
[270]
[271]
[272]
[273]
[274]
[275]
[276]
[277]
[278]
[279]
[280]
[281]
[282]
[283]
[284]
[285]
[286]
[287]
[288]
[289]
[290]
[291]
[292]
[293]
[294]
[295]
[296]
[297]
[298]
[299]
[300]
[301]
[302]
[303]
[304]
[305]
[306]
[307]
[308]
[309]
[310]
[311]
[312]
[313]
[314]
[315]
[316]
[317]
[318]
[319]
[320]
[321]
[322]
[323]
[324]
[325]
[326]
[327]
[328]
[329]
[330]
[331]
[332]
[333]
[334]
[335]
[336]
[337]
[338]
[339]
[340]
[341]
[342]
[343]
[344]
[345]
[346]
[347]
[348]
[349]
[350]
[351]
[352]
[353]
[354]
[355]
[356]
[357]
[358]
[359]
[360]
[361]
[362]
[363]
[364]
[365]
[366]
[367]
[368]
[369]
[370]
[371]
[372]
[373]
[374]
[375]
[376]
[377]
[378]
[379]
[380]
[381]
[382]
[383]
[384]
[385]
[386]
[387]
[388]
[389]
[390]
[391]
[392]
[393]
[394]
[395]
[396]
[397]
[398]
[399]
[400]
[401]
[402]
[403]
[404]
[405]
[406]
[407]
[408]
[409]
[410]
[411]
[412]
[413]
[414]
[415]
[416]
[417]
[418]
[419]
[420]
[421]
[422]
[423]
[424]
[425]
[426]
[427]
[428]
[429]
[430]
[431]
[432]
[433]
[434]
[435]
[436]
[437]
[438]
[439]
[440]
[441]
[442]
[443]
[444]
[445]
[446]
[447]
[448]
[449]
[450]
[451]
[452]
[453]
[454]
[455]
[456]
[457]
[458]
[459]
[460]
[461]
[462]
[463]
[464]
[465]
[466]
[467]
[468]
[469]
[470]
[471]
[472]
[473]
[474]
[475]
[476]
[477]
[478]
[479]
[480]
[481]
[482]
[483]
[484]
[485]
[486]
[487]
[488]
[489]
[490]
[491]
[492]
[493]
[494]
[495]
[496]
[497]
[498]
[499]
[500]
[501]
[502]
[503]
[504]
[505]
[506]
[507]
[508]
[509]
[510]
[511]
[512]
[513]
[514]
[515]
[516]
[517]
[518]
[519]
[520]
[521]
[522]
[523]
[524]
[525]
[526]
[527]
[528]
[529]
[530]
[531]
[532]
[533]
[534]
[535]
[536]
[537]
[538]
[539]
[540]
[541]
[542]
[543]
[544]
[545]
[546]
[547]
[548]
[549]
[550]
[551]
[552]
[553]
[554]
[555]
[556]
[557]
[558]
[559]
[560]
[561]
[562]
[563]
[564]
[565]
[566]
[567]
[568]
[569]
[570]
[571]
[572]
[573]
[574]
[575]
[576]
[577]
[578]
[579]
[580]
[581]
[582]
[583]
[584]
[585]
[586]
[587]
[588]
[589]
[590]
[591]
[592]
[593]
[594]
[595]
[596]
[597]
[598]
[599]
[600]
[601]
[602]
[603]
[604]
[605]
[606]
[607]
[608]
[609]
[610]
[611]
[612]
[613]
[614]
[615]
[616]
[617]
[618]
[619]
[620]
[621]
[622]
[623]
[624]
[625]
[626]
[627]
[628]
[629]
[630]
[631]
[632]
[633]
[634]
[635]
[636]
[637]
[638]
[639]
[640]
[641]
[642]
[643]
[644]
[645]
[646]
[647]
[648]
[649]
[650]
[651]
[652]
[653]
[654]
[655]
[656]
[657]
[658]
[659]
[660]
[661]
[662]
[663]
[664]
[665]
[666]
[667]
[668]
[669]
[670]
[671]
[672]
[673]
[674]
[675]
[676]
[677]
[678]
[679]
[680]
[681]
[682]
[683]
[684]
[685]
[686]
[687]
[688]
[689]
[690]
[691]
[692]
[693]
[694]
[695]
[696]
[697]
[698]
[699]
[700]
[701]
[702]
[703]
[704]
[705]
[706]
[707]
[708]
[709]
[710]
[711]
[712]
[713]
[714]
[715]
[716]
[717]
[718]
[719]
[720]
[721]
[722]
[723]
[724]
[725]
[726]
[727]
[728]
[729]
[730]
[731]
[732]
[733]
[734]
[735]
[736]
[737]
[738]
[739]
[740]
[741]
[742]
[743]
[744]
[745]
[746]
[747]
[748]
[749]
[750]
[751]
[752]
[753]
[754]
[755]
[756]
[757]
[758]
[759]
[760]
[761]
[762]
[763]
[764]
[765]
[766]
[767]
[768]
[769]
[770]
[771]
[772]
[773]
[774]
[775]
[776]
[777]
[778]
[779]
[780]
[781]
[782]
[783]
[784]
[785]
[786]
[787]
[788]
[789]
[790]
[791]
[792]
[793]
[794]
[795]
[796]
[797]
[798]
[799]
[800]
[801]
[802]
[803]
[804]
[805]
[806]
[807]
[808]
[809]
[810]
[811]
[812]
[813]
[814]
[815]
[816]
[817]
[818]
[819]
[820]
[821]
[822]
[823]
[824]
[825]
[826]
[827]
[828]
[829]
[830]
[831]
[832]
[833]
[834]
[835]
[836]
[837]
[838]
[839]
[840]
[841]
[842]
[843]
[844]
[845]
[846]
[847]
[848]
[849]
[850]
[851]
[852]
[853]
[854]
[855]
[856]
[857]
[858]
[859]
[860]
[861]
[862]
[863]
[864]
[865]
[866]
[867]
[868]
[869]
[870]
[871]
[872]
[873]
[874]
[875]
[876]
[877]
[878]
[879]
[880]
[881]
[882]
[883]
[884]
[885]
[886]
[887]
[888]
[889]
[890]
[891]
[892]
[893]
[894]
[895]
[896]
[897]
[898]
[899]
[900]
[901]
[902]
[903]
[904]
[905]
[906]
[907]
[908]
[909]
[910]
[911]
[912]
[913]
[914]
[915]
[916]
[917]
[918]
[919]
[920]
[921]
[922]
[923]
[924]
[925]
[926]
[927]
[928]
[929]
[930]
[931]
[932]
[933]
[934]
[935]
[936]
[937]
[938]
[939]
[940]
[941]
[942]
[943]
[944]
[945]
[946]
[947]
[948]
[949]
[950]
[951]
[952]
[953]
[954]
[955]
[956]
[957]
[958]
[959]
[960]
[961]
[962]
[963]
[964]
[965]
[966]
[967]
[968]
[969]
[970]
[971]
[972]
[973]
[974]
[975]
[976]
[977]
[978]
[979]
[980]
[981]
[982]
[983]
[984]
[985]
[986]
[987]
[988]
[989]
[990]
[991]
[992]
[993]
[994]
[995]
[996]
[997]
[998]
[999]
[1000]

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF102855; AAD04569.2; --
EMBL; AF131951; AAD29417.1; ALT_INIT.
EMBL; AF159046; AAD42975.1; --
EMBL; AF141904; AAF02498.1; ALT_INIT.
HSSP; P00519; 1ABL.
InterPro; IPR002110; ANK.
InterPro; IPR001478; PDZ.
InterPro; IPR001660; SAM.
InterPro; IPR001452; SH3.
Pfam; PF00023; ank; 7.
Pfam; PF00595; PDZ; 1.
Pfam; PF00536; SAM; 1.
Pfam; PF00018; SH3; 1.
ProDom; PD000066; SH3; 1.
SMART; SM00248; ANK; 6.
SMART; SM00228; PDZ; 1.
SMART; SM00454; SAM; 1.
SMART; SM00326; SH3; 1.
PROSITE; PS50088; ANK_REPEAT; 3.
PROSITE; PS50297; ANK_REPEAT; 1.
PROSITE; PS50106; PDZ; 1.
PROSITE; PS50002; SH3; 1.
PROSITE; PS50105; SAM_DOMAIN; 1.
ANK repeat; SH3 domain; Repeat; Alternative splicing.
REPEAT 195 210 ANK 1.
REPEAT 212 245 ANK 2.
REPEAT 246 278 ANK 3.
REPEAT 279 312 ANK 4.
REPEAT 313 345 ANK 5.
REPEAT 346 378 ANK 6.
REPEAT 379 395 ANK 7.
DOMAIN 554 613 SH3.
DOMAIN 663 757 PDZ.
DOMAIN 2104 2167 SAM.
POLY-PRO.
POLY-HIS.
POLY-HIS.
POLY-GLY.
POLY-PRO.
Missing (in isoform 2).
/FTid=VSP_006072.
SQEGRQSRSDKAKRLFRHYTVGYDSFDAPSLDIDSG
-> MALSAVGCGPGGALPQPFPALSSSWPAGPRRSVWY
IY (in isoform 2).
/FTid=VSP_006073.
Missing (in isoform 3).
/FTid=VSP_006074.
Missing (in isoform 4).
/FTid=VSP_006075.
LSEDSQTSLSKPS -> QYRIVVKSSDFGDF (in isoform 5).
/FTid=VSP_006076.
Missing (in isoform 5).
/FTid=VSP_006077.
S -> T (IN REF. 1).
S -> N (IN REF. 2).
R -> K (IN REF. 1).
A -> T (IN REF. 1).
S -> D (IN REF. 1).
S -> N (IN REF. 2).
3.3%; Score 432; DB 1; Length 2167;
Best Local Similarity 21.0%; Pred. No. 4.7e-07;
Matches 394; Conservative 139; Mismatches 619; Indels 722; Gaps 87;
Query Match
Sequence 2167 AA; 226333 MW; 3F478B5A7H18BA86 CRC64;

RA Angrand P.O., Valvatne H., Jeanmougin F., Adamson A.,
RA van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,
RA Lamerdin J., Chambon P., Lossen P., Stewart A., Aasland R.,
RT "Mammalian trithorax- and ASH1-like proteins: putative chromatin
RT regulators which contain PHD fingers and SET domains";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Lamerdin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,
RA Garcia E., Kyle A., Ramirez M., Stilwagen S., Garges J., Danganan L.,
RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,
RA Carrano A.V.;
RT "Sequence analysis of a 1 Mb region in human 19q13.1";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 111-2715 FROM N.A. (ISOFORM 1).
RC TISSUE=Leukocyte, and Testis;
RX MEDLINE=20105772; PubMed=10637508;
RA Hunsman D.G., Chin S.-F., Mulleris M., Batley S.J., Collins V.P.,
RA Wiedemann L.M., Aparicio S., Caldas C.;
RT "MLL2, the second human homolog of the Drosophila trithorax gene, maps
RT to 19q13.1 and is amplified in solid tumor cell lines";
RL Oncogene 18:7975-7984(1999).
RN [4]
RP SEQUENCE OF 816-2715 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205941;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro";
RL DNA Res. 4:141-150(1997).
RN [5]
RP SEQUENCE OF 1918-2715 FROM N.A.
RC TISSUE=Brain, and Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting M., Madan A., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Bone marrow, and Placenta;
RX MEDLINE=99339983; PubMed=10409430;
RA FitzGerald K.T., Diaz M.O.;
RT "MLL2: A new mammalian member of the trx/MLL family of genes";
RL Genomics 59:187-192(1999).
CC -1- FUNCTION: Possibly acts as a transcriptional regulatory factor.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=Long;
CC IsoId=Q9UNW6-1; Sequence=Displayed;
CC Name=2; Synonyms=Truncated; Sequence=VSP_006668;
CC IsoId=Q9UNW6-2; Sequence=VSP_006668, VSP_006669;
CC -1- TISSUE SPECIFICITY: Widely expressed. Highest levels in testis.

CC Also found in brain, bone marrow, heart, muscle, kidney, pancreas,
CC spleen, thymus, prostate, ovary, intestine, colon, peripheral
CC blood lymphocytes, and placenta.
CC -1- DISEASE: Often amplified in pancreatic carcinomas.
CC -1- SIMILARITY: Belongs to the TRX/MLL family.
CC -1- SIMILARITY: Contains 1 bromodomain.
CC -1- SIMILARITY: Contains 3 PHD-type zinc fingers.
CC -1- SIMILARITY: Contains 1 CXXC-type zinc finger.
CC -1- SIMILARITY: Contains 1 post-SET domain.
CC -1- CAUTION: This protein was first named MLL2 by Ref.3 and Ref.6.
CC MLL2 corresponds to another protein located on chromosome 12 (see
CC AC 014686).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ007041; CAB45385.1; -;
CC EMBL; AD000671; -; NOT_ANNOTATED_CDS.
CC EMBL; AB002302; BAA20763.2; -;
CC EMBL; AF186605; AAD56420.1; -;
CC EMBL; AF104918; AAD17932.1; -;
CC EMBL; AF105279; AAD26113.1; -;
CC EMBL; BC009337; AAH09337.1; -;
CC EMBL; BC007353; AAH07353.1; -;
CC EMBL; AF105280; AAD26112.1; -;
CC MIM; 606834; -;
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0003700; F:transcription factor activity; NAS.
CC GO; GO:0008270; F:zinc ion binding; NAS.
CC GO; GO:0048096; P:chromatin-mediated maintenance of transcrip. . .; NAS.
CC InterPro; IPR003889; FYRICH_C.
CC InterPro; IPR003616; PostSET.
CC InterPro; IPR001214; SET.
CC InterPro; IPR02857; Znf_CXXC.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF00628; PHD; 3.
CC Pfam; PF00856; SET; 1.
CC Pfam; PF02008; zf-CXXC; 1.
CC SMART; SM00542; FYRC; 1.
CC SMART; SM00541; FYRN; 1.
CC SMART; SM00249; PHD; 4.
CC SMART; SM00508; PostSET; 1.
CC SMART; SM00317; SET; 1.
CC PROSITE; PS50868; POST_SET; 1.
CC PROSITE; PS50280; SET_1.
CC PROSITE; PS01359; ZF_PHD_1; 3.
CC PROSITE; PS50016; ZF_PHD_2; 3.
CC DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;
CC Transcription regulation; Alternative splicing; Repeat.
CC A.T HOOK (BY SIMILARITY).
CC DNA_BIND 37 44
CC FT 110 117
CC FT DNAS_BIND 357 365
CC FT DNAS_BIND 959 1005
CC FT ZN_FING 1201 1252
CC FT ZN_FING 1249 1303
CC FT ZN_FING 1335 1396
CC FT ZN_FING 1449 1471
CC FT DOMAIN 2574 2695
CC FT DOMAIN 2699 2715
CC FT DOMAIN 26 37
CC FT DOMAIN 248 255
CC FT DOMAIN 362 398
CC FT DOMAIN 402 771
CC FT DOMAIN 808 812
CC FT DOMAIN 1963 1970
CC FT DOMAIN 2251 2259
CC FT DOMAIN 1963 1970
CC FT DOMAIN 2251 2259

FT	VARSPLIC	532	582	VSARSRIKTPRRWDESDPPKPPKVEVSPVLRPPITSP	
FT	FT			VQEEAPVPS -> PLSQSLPLPMTQLSLGQWAAPTTS	
FT	FT			ACLDSPMLRLLRRCPLTLQL (in isoform 2).	
FT	FT	583	2715	Missing (in isoform 2).	
FT	FT			/FTid=VSP 006668.	
FT	FT			/FTid=VSP 006669.	
FT	CONFLICT	834	834	K -> E (IN REF. 6).	
FT	CONFLICT	941	941	S -> Y (IN REF. 6).	
FT	CONFLICT	1317	1317	E -> Q (IN REF. 6).	
FT	CONFLICT	1362	1362	H -> Y (IN REF. 6).	
FT	CONFLICT	1438	1438	D -> N (IN REF. 6).	
FT	CONFLICT	1918	1920	PLA -> GTR (IN REF. 5; AAH09337).	
FT	CONFLICT	2541	2543	DEE -> ARG (IN REF. 5; AAH07353).	
FT	CONFLICT	2622	2622	D -> H (IN REF. 6).	
FT	SEQUENCE	2715	AA; 293511	MW; C0615B981BEB7BF CRC64;	
Query Match					3.2%; Score 428.5; DB 1; Length 2715;
Best Local Similarity					19.0%; Pred. No. 7.5e-07;
Matches 515; Conservative 223; Mismatches 798; Indels 1171; Gaps 124;					
Qy	488	RRRKSQQQQQQQ	-----QQQQQQQQMPRSSQBDEKEKEK	-----	527
Db	179	KRGEGEGRVWQALTELLRRAQAQAPRSRACEPSTPRSRGRPPGRPAGCRKQQAQV	238		
Qy	528	---EA-----EKEEKPFVENDKEDL--LKEKTDTSGEDNDEKEA VASKRKTANSQGR	578		
Db	239	VVAEAAVTIPKEPPPPVVPVKHGTGSKWCKEGPGPGCTPRRGQSRGGRGGRG	298		
Qy	579	KGR-----ITRSMANEANSEAI TPQSAELASLMELESSRWTEEMETAKGLLEHGR	632		
Db	299	GGLPFVIKVSRA-----KKVXWGQSLGLES-----GQGGQHEE	334		
Qy	633	NWSAI-ARMVGS-KTVSCKNFYFNKQKQNLDEILOQHKLMKERNARKKKA	686		
Db	335	SMQDVPQVRVSGQSGSPWK-----KQEKLDDEEEKKEEKQKKEGEKEKRAVAEE	389		
Qy	687	---PAAASEEAPFPVVEDEEASGVSGNEBEMVEEAEALHASGNEVPRGCSGPATVNN	744		
Db	390	NMPAAEKEEAKLPP-----	403		
Qy	745	SSDTESIPSPHTEAAKDTGQNGKPPATLGADGPPGPPPTPRRTSRAPISBTPASEATG	804		
Db	404	-----PPLTPAPSP--PPLPPPTS	423		
Qy	805	APTTP--PAPPSGAPPVPVKEEKEETAAPVVEEGEOKPPAAAEELAVDTGKAERP	862		
Db	424	-PPPLCPPPPPVSPPL-----PSPPPPAQEQQEESPVPVVPATCSRKGRPPL	474		
Qy	863	KSECTEAEAGPAKGDABAEATAEGALKAEKKGSGRATTAKSGAPQDSSATCS	922		
Db	475	T-----PSQAREEAAARAGPETSPTTP-----TPSTATGGPPEDSPVAPKS	517		
Qy	923	ADEVEAEGGDKN-----RLLSRPPLTLTGDPANASQKPLDLK---OLQORAAAI	973		
Db	518	TTFL-----KNIRQFIMPVSARSRVIKTPRRFMDEDPKPKPEVSPVLRPPITTS	570		
Qy	974	PPi--QVTKVHEPPREDAAPTAPAPAP-----PPQNLOPE	1008		
Db	571	PPVQEPAPVSPR--APTPTPTVLPKERSILREPTFRWTSLTRPLPPPPPPPP	627		
Qy	1009	SDAFQPGSGSRKSRSP-----AP---PADKEAFAAEAKLPGDPPCWTSGLPFPVPPRE	1061		
Db	628	PPAPSP--PATATSSRRLLLRAPQFTPSEAHKIYESVLTP--PPLGAPEAPEPEP--	682		
Qy	1062	VIKASPHAPDPSARYAPPGLPLGLHDTARVLP--PTISNPPPLISS	1111		
Db	683	PADDSAPAEPEPRAV-----GRTNHLSLPRFA-PVVTPVKAESVPHGAPALSNPGPQTAQ	736		
Qy	1112	AKHPSVLERIGATISQGSVOLHVPYSEHAKAPVPTMGLPLMDPKLA-----PFSG	1166		
Db	737	LLQP-----LQALQTQLLPALPPQPLQPPSPQOMP---PLEKARIAGVCSLPLSG	787		
Qy	1167	VKBEQ-----LSPRGQ-----	1177		

Db	788	VEEKNFSLKRAKVLQFKIDQOQQQKVAASMPLSFGQMEEVAGAVKQISDRGPVSEDE	847		
Qy	1178	-----AGPESLGV--TAQASVLRTALGVPGSGITKGPSTVRPDSAIT	1224		
Db	848	SVEAKRPSGSPVQGPRIKHVCRHAVALQARAWP-----EDVPLSALP	897		
Qy	1225	YRGSITHGTADVLYKGTITRIIGEDSPSLDRGRED-----	1261		
Db	898	LRDR-----QLATEDTSSASETESVFSRRRGKVEAAGPGESEPTSGGTLAHTPR	951		
Qy	1262	SLPKGHVIEGKK-----GHVLSYEGGMSTQCSK-----EDGRSSSP-PHETAAPKRT	1310		
Db	952	SLPSHH-----GKQMRMARGHC-----RGLRVQDCGVCVNCCLKPKFGPNTKKQCCVVRK	1004		
Qy	1311	YDMEGR-----VGRAI-----SSASIEGLMGRAIPEERHSPHLLKEQHIRGSI	1355		
Db	1005	CDKTEARWELAKGRITVITLLPWDSDESPASG--PP-----GPR-----RGAG	1050		
Qy	1356	TQGIPRSYV-----EAQEDYLREAA-----	1380		
Db	1051	AGG-PREEVVAHPGPEQDSLLQKSAARRCVKQPSYDIFEDSDDEPGPPAPRRTPR	1109		
Qy	1381	EGTPPPPPPSR-----	1391		
Db	1110	ENELPLPEPERQSRPRKPTLQVLQKARRRLDKDALAPGPFASPPNGWTGKQKSPGVH	1169		
Qy	1392	-----DLTEAYKTQALGPLKPKAHEGLVATVKEAGRSIHEIPREELRHTPELPL	1441		
Db	1170	RVRVDFKEDCLENVLMGLSVLTSVPGPPMVCLLC-ASKGLHELVCQVCCDFHFP	1228		
Qy	1442	---APRPLEKSGITQGTPLKYDT-----GASTTGSK-----	1469		
Db	1229	CLEEAERPL-----POHHDWCCRCCKFCHVCKGRGSKHLLBECERCHAYHPAC	1279		
Qy	1470	-----KHDV-----RSLIGSPGRTPFPVHPLDVMADARALE-----	1500		
Db	1280	LGPSYPTRATRRRHVICSAVCRCSCGATGKNWVWSGDSYLCRCTLQYKGNVCP	1339		
Qy	1501	---RACYE-----ESLKSRCPTASSSGSIARGA--	1526		
Db	1340	ICTRCYEDNDYESKMMQCAQCDHWHAKEGSLDEDEYELSLGPSVLVTCGPCAGAAQ	1399		
Qy	1527	-----PVIVELGKPROSPLTYEDHGAPFAGHLRPGSPVTWREPTRLQE	1571		
Db	1400	RWREALSGALQGLRQVLQGLSSKVVGPLLLCTCGDQDGLQHPG-FCGLQAVSORFED	1458		
Qy	1572	GSLSKASQDRKLTSTPREIAKSPHSTVPE-----HHPHPISPYEH--LRGV---	1618		
Db	1459	GHY-----KSVHSPMEDVMVGLMRHSEGETPDRRAGQGMKGLLK	1499		
Qy	1619	---SGVDLYRSHIPLAF-DPTSIPRGIPLDAAYYLPRLHLPNPTYPHLY-----	1665		
Db	1500	LLESFAFGFADPKYWRRTLNGV-----LPNAVLP-PSLDHVAQMRQBPPE	1549		
Qy	1666	---PPYLIRGYP-----DTAAL-----ENR	1682		
Db	1550	TPESQPP-----GDSAAFOGKDPAAASHLEDPQCALCKYGDADSKAEGRLLYIGNE	1605		
Qy	1683	QTIINDYITSOQMH-----NTATAMAORADM-----	1709		
Db	1606	WTHVNCATWSAEVPEENDSLKNVHAARQMRQRCCLCKPGATVGCCLSSCLSNFHM	1665		
Qy	1710	-----	1715		
Db	1666	CARASYCIFODDKVKVCKQKHTDLDGKEI VNPQDGVLLRVYVDFEGINFKRKLTLGLEP	1725		
Qy	1716	RESSIALNYAAGPGIIDLQSOVPHL-----PV-----LV	1744		
Db	1726	D-----AINVLIGSTRIDSLGTLSDCEGLRFIGYQCSRLYSTVDARRCWRRCIL	1781		
Qy	1745	PPTGTPATAMDRLAYLPTAPQPPSSRHSPLS--PGGPHLTHTKPTTTSSSRRERDR	1802		

Db 1782 EYRPGW---REEPAHLEAAEENQTIIVHSPAPSEPPGG----- 1817
Qy 1803 ERDRERREKILSTTTTVEHAPLWR---PGTEQSSGSSGGGGSSRRSPASHAHQH 1860
Db 1818 ---EDPPLDITDVLVPGAPERHSPIONIDPLLRPDGSA----- 1852
Qy 1861 SPISPRTDALQORPSVLHNTGMGIITAVEPSPKPTVLRLSTSTSSPVVRPAATPPATHCP 1920
Db 1853 PPPAPRS-----FSGARIKVPNTYSPRR-----P 1876
Qy 1921 LGGTLDGYPTLMBPVLLPKPEAPVARPERPRADTGHAFIAKPPAR-----SGLEPASPPS 1976
Db 1877 LGGVSGFPLPSPGFSLLTHIPTVGPDPF-----APRRRRRPSPLAPRPPS 1926
Qy 1977 KGSEPRLL-----VPPVSGHATITAPKNIAPHASDPDPAP----- 2014
Db 1927 RWASP-PLKTSPPQURVPPPTSVVLTPTSGELAPPAPSPPPEDLGDFFEDMEVUSG 1985
Qy 2015 -PASASDPHREKTKSPFSIQEELSLRSLGVHSGSYSPGVE-----PVSPV 2059
Db 1986 LSAADLFAASLLGTEFP--QE-EIVAAGAMGSSHGSGGDSSEESSPTSRYIHFPVTVV 2042
Qy 2060 SSPSL--THDKGLPKHLELD-----KSHLEGELPK-----QPGPVKLGGEAAHLP 2104
Db 2043 SAPGLAPSATGAPR-IEQLDGVDDGTDEAEAVQPRGQGTTPSPGPGVVRAGVLGAAGD 2101
Qy 2105 HLRPLPESQSSPLLQTPAGVKHQHVVTTLAQHISEVITDYYTRHHPOQLSAPLPAPLY 2164
Db 2102 RARP-PEDLP-----SEIV--DFVL-----KNLGGP----- 2124
Qy 2165 SFGASCPCVLDLRPPPSDLYLPDPDHGAP--ARGSPHSEGGKSP-BPNKTSVLGGGEGD 2221
Db 2125 GDGGAG-----PREESLPP---APPLANGSQPSQGLTASPADPTRTFAMLPAGP 2171
Qy 2222 IEPVS-----PPEGMTPE--GHSRAVYPLLVRDGEQTPSRMGSKSPGNTSQPPAF 2272
Db 2172 VRVLSLGPAPPPKPPATSKILVNKLQGVFVMAGEGEPVPPP-----VKQPL-- 2220
Qy 2273 SKLTESAMVKKQKQKINKKLNTHNRNEPYNISQCPTEIFNMPAITGTGLMTYRSQAV 2332
Db 2221 -----PPIISPTAPTSWILPQGLLVLPV----- 2246
Qy 2333 QEHASTNMGLEAIRKA-----LMGKYDQWEESEPLSA-----NAFNPLNASALPA 2379
Db 2247 -----GVVRPAPPPPPPLTLVLSSGPASPPRQAIRVVRKSVTFSGKSPAPPPY 2295
Qy 2380 AMPITAADGR-SDHTLTSPGGGKAKVSCRPSRRKAKSP-APGLASGDRPPSVSVHSEG 2437
Db 2296 KAPRLDEDEGEASEDTPQVPLG-----SGGFSRVRMKTPTVRGVLDLDRPGEPAGEESPG 2350
Qy 2438 DCNRRTPLTNRWEDRPSAGSTPPYNNPLIMRLQ-----AGVMAS-----PPPPGLPAGSG 2489
Db 2351 FLQERSPLL-PLPEDGPPQVDPGP---PDLLLESQWHHSYGEASSSEEPSPDDKENQA 2406
Qy 2490 P-LAGPH 2495
Db 2407 PKRTGPH 2413

RESULT 21
NC06 HUMAN
ID NC06 HUMAN STANDARD; PRT: 2063 AA.
AC Q14686; Q9UNT9; Q9UH74; Q9UK86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nuclear receptor coactivator 6 (Amplified in breast cancer-3 protein)
DE (Cancer-amplified transcriptional coactivator ASC-2) (Activating
DE signal cointegrator-2) (ASC-2) (Peroxisome proliferator-activated
DE receptor-interacting protein) (PPAR-interacting protein) (PRIP)
DE (Nuclear receptor-activating protein, 250 kDa) (Nuclear receptor
DE coactivator RAP250) (NRC RAP250) (Thyroid hormone receptor-binding
DE protein).

GN NCOA6 OR AIB3 OR RAP250 OR TRBP OR KIAA0181.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND INTERACTION WITH CREBBP; NCOA1; GTF2A; TBP;
RX RXRA; ESR1; RARA AND THRA.
RA Lee S.-K., Anzick S.L., Choi J.-E., Bubendorf L., Guan X.-Y.,
RA Jung Y.-X., Kallioniemi O.P., Kononen J., Trent J.M., Azorsa D.,
RA Jhun B.-H., Cheong J.H., Lee Y.C., Meltzer P.S., Lee J.W.;
RT "A nuclear factor ASC-2, as a cancer-amplified transcriptional
RT coactivator essential for ligand-dependent transactivation by nuclear
RL receptors in vivo."; J. Biol. Chem. 274:34283-34293 (1999).
RN [2]
RP SEQUENCE FROM N.A., HOMODIMERIZATION, AND INTERACTION WITH CREBBP;
RX RXRA; ESR1; NR3C1; RARA; VDR AND THRA.
RA MEDLINE=20036574; PubMed=10567404;
RA MEDLINE=20325329; PubMed=10866662;
RA Mahajan M.A., Samuels H.H.;
RT "A new family of nuclear receptor coregulators that integrates nuclear
RT receptor signaling through CBP."; Mol. Cell. Biol. 20:5048-5063 (2000).
RN [3]
RP SEQUENCE FROM N.A., AND INTERACTION WITH PPARA; PPARG; ESR1; ESR2 AND
RX THRA.
RA TISSUE=Testis; PubMed=10681503;
RA MEDLINE=20148724; PubMed=10681503;
RA Cairns F., Antonson P., Pelto-Huikko M., Treuter E., Gustafsson J.-A.;
RT "Cloning and characterization of RAP250, a nuclear receptor
RT coactivator."; J. Biol. Chem. 275:5308-5317 (2000).
RN [4]
RP SEQUENCE FROM N.A., PHOSPHORYLATION BY PRKDC, AND INTERACTION WITH
RX THR; RAR; EP300 AND CRSP3.
RA TISSUE=Lymphocytes; PubMed=10823961;
RA MEDLINE=20283976; PubMed=10823961;
RA KO L., Cardona G.R., Chin W.W.;
RT "Thyroid hormone receptor-binding protein, an LXXLL motif-containing
RT protein, functions as a general coactivator."; Proc. Natl. Acad. Sci. U.S.A. 97:6212-6217 (2000).
RN [5]
RP SEQUENCE FROM N.A.
RX TISSUE=Bone marrow; PubMed=8724849;
RA MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RL analysis of cDNA clones from human cell line KG-1."; DNA Res. 3:17-24 (1996).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.B.,
RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., Fraser L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehtvasilao M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McClay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Rameay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,

Db	1356	-----VSEFSEAKDENERSIASLSPMDPEVPDSESPVE-KVLSPLRSPPLLGSESPY--	1405
Qy	1366	AQEDYLREAKLKR-----EGTTPPPPSRDLTAYKTCALGPLKLPKPAHEG	1413
Db	1406	--EDFLSADSKVLGRRESPEFGKNGKQGFDRSPVSDUT-----STGLYDQKEKSTG	1459
Qy	1414	LVAITKEAGRSIHEIPREELRHTEPLAPRLPKEGSITQCTPLKYDTGASTTGSKKHDV	1473
Db	1460	FIPKEDGPEBKTSVDVETWSSQSALADERKL-GGDV-----SPTQID--VSQFGSPKEDT	1513
Qy	1474	RLSIGSPORTPPVPHPLDVNADARALERACYVESLKRPGTASSGSGSIARGAPVIVPEL	1533
Db	1514	KWSISEGTVSKSATPD-----EGVAEDTSHMEGVASVSTASVATSS---FPPEP	1561
Qy	1534	GKQPSPLTYDHGAPAGHLPGRSPVTMREPTRLQEGSLSSSKASQDRKLTSTPREI-	1592
Db	1562	TTDDVPSPLHAEGVSPHSTEVDDSLSVSVVQTPTTFQETEMSPSKCECPRPMISPPDFS	1621
Qy	1593	--AKSPHSTVPEHHPHPISPYEHLLRGVGDVLYRSHIPLAFDPTSPRGIPLDAAAAY	1649
Db	1622	PKTAKS-RTPVQDRHSE-----QSSMSIEFGQESPEHSFAMDFS---1659	
Qy	1650	YLPRLHLPNTPYPHLYPPLRGVPTDRALENQTIINDYITSQMHNTATAMAQRADM	1709
Db	1660	-----RQSPDHPTL-----GASVLHITENGTEV-DY-----1685	
Qy	1710	LRGLSP---RESSIALNVAAGRGIIIDLSQVPHLPVLPVPPPGTATAMDRLAYLTPAQ	1766
Db	1686	-----SPCDIQDSSLHK-----IPTEEPSYIQDNDLSLISVSQ	1721
Qy	1767	-----PFSRHS-----SSPLSPGGTHLTKPTTS-----SSERDRDRDRDREREK	1812
Db	1722	VEASPSTSSAHTPSQIAPLOBDTLSDVVPPREMSLYASLASEKVSQLEGE-----KLSPK	1777
Qy	1813	SILTSTTVEHAPIWRPQTEQSSGSGSGSGGSRPASHAHQSPISPRQTQALQ	1872
Db	1778	SDISPLTPRESSPLSPGSDSTAETAAMAHQASSP-----PIDATABPYG	1827
Qy	1873	QRPVLHNTGMKGIITAVEPSKPTVLRSTSTSSVVRPAATPPATHCPGLGTLGCVPTL	1932
Db	1828	FRSSMLFDTMQHHL-----ALNRDLTSSVEKSGCKTP-----GDFNYAYQKP	1871
Qy	1933	MEVLLPKBA-----PRVAPERPRADTGHAF-----LAKPPAR	1966
Db	1872	ENAAAGSPDEEDYDSEKTIIRHDVVRYVEKTEKTKSPCDSGSYETIEKTKITPED	1931
Qy	1967	SGL-----EPASSPSKG-----SEPRPLVPPVSGHATARTAPKNLAPH-----	2005
Db	1932	GGYTCTITKTTPTEGGYSVEISEKTTTPEVSGTYVEKTERSRLLDDISNGYDDTE	1991
Qy	2006	---HASFPDPAPPASADPHREKTSQKPSIQEILSRSLGYH-----GSSYSPEG	2052
Db	1992	DGCGHTLGD-----CSYSYETTEKITSP-----ESESYSYETSTKTSRPTSAYCYET	2040
Qy	2053	VEPVSPVSSPS-----LTHDKGLPKH-----LBELDKSHLEGLRPK--QP	2091
Db	2041	MEKITTPQASTVSYTSDRCYTEKKSSEARQVDVLCVSSCEFKHFKTELSPSFNP	2100
Qy	2092	GPVK-LGGEAAHLPLRLPESQPSPLQOT--APGVKHQRVWTLAHISEVITQDYT	2148
Db	2101	NPLEWFAGE-----EPTSESE---KPLTQSGGAPPPSGGKO---QGRQCDTEPTTSVS	2147
Qy	2149	RHHQQLSALPAPLPSFPFGASCPLV-----DLRRPPSDLYL-----PPPDHGA	2192
Db	2148	ESAPSQDSDVPPPT-----EECPSTADANIDSESESETIPTDKVTYKHMDDPP--A	2199
Qy	2193	PARGSPHSEGGKRPENKTSVLGGGBDGI-----EPVSPPEGMTEPCHERSAV	2241
Db	2200	PMQ-----DRSPSRHPDVSVMVDPDALAVQNLGKAVKDKLKEKTKTKKSGTKSS	2251
Qy	2242	YPLLYRGEQTEPBRMGSKSPGNTSQPPAFPSKLTESAMVSKKQKQKINKLNTHRNE	2301

Db	2252	SPVKKGDGK-----SKPLAASPFGALKESSDKSVRSVPKPKKESVEKA-TKTTTT	2301
Qy	2302	PEYNIQSQTGTFIMPAITGTLMTYRSQAVOEHAFTNMGLEAIRKALMGKYDQWEESE	2361
Db	2302	PEV-----KATRG-----BEKD	2313
Qy	2362	PLSANAFNPLNASLPAAMETITADGRSDHTLTPSGGGGKARVSGRPSSRKAAPGL	2421
Db	2314	KETQNA-----ANASAKSAK---TAT-----TGPGTTTKAKSTVDP-----PGL	2350

RESULT 23

ID	SNT1	YEAST	STANDARD;	PRT;	1226	AA.
AC	P25357;	Q02397;	Q8N1L8;			
DT	10-MAY-1992	(Rel. 22, Created)				
DT	10-OCT-2003	(Rel. 42, Last sequence update)				
DE	Potential DNA-binding protein SNT1.					
GN	SNT1 OR YCR033W OR YCR33W OR YCR592.					
OS	Saccharomyces cerevisiae (Baker's Yeast).					
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;					
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.					
OX	NCBI_TaxID=4932;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=92133156; PubMed=1776366;					
RA	Wicksteed B.L., Roberts A.B., Sagliocco F.A., Brown A.J.P.;					
RT	"The complete sequence of a 7.5 kb region of chromosome III from					
RT	Saccharomyces cerevisiae that lies between CRY1 and MAT.";					
RL	Yeast 7:761-772(1991).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=91335897; PubMed=1872032;					
RA	Jia Y., Slonimski P.P., Herbert C.J.;					
RT	"The complete sequence of the unit YCR59, situated between CRY1 and					
RT	MAT, reveals two long open reading frames, which cover 91% of the					
RT	10.1 kb segment.";					
RL	Yeast 7:413-424(1991).					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=S288C;					
RX	MEDLINE=92244356; PubMed=1574125;					
RA	Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,					
RA	Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,					
RA	Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,					
RA	Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,					
RA	Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,					
RA	De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,					
RA	Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,					
RA	Fairhead C.A., Faye G., Feldmann H., Fiers W., Frontali L., Fukuhara H.,					
RA	Franciques-Gaillard M.-C., Franco L., Gilliquet V., Glandsdorff N.,					
RA	Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Grivell L.A., Haesemann M.,					
RA	Goffeau A., Gresson M., Grisanti P., Gruber C.J., Hilger F., Hohmann S.,					
RA	Hatad D., Hegemann J.H., Herbert C.J., Indge K.J., Isono K., Jackson P.,					
RA	Hollenberg C.P., Huse K., Iborra F., Jaumaux J.-C., Jia Y., Jimenez A.,					
RA	Jaou C., Jacquet M., James C.M., Jaumaux J.-C., Lewis C., van der Linden C.G.,					
RA	Kleinhaus U., Kreisl P., Lafranchi G., Lewis C., Maat C., Mannhaupt G., Manzano M.E.,					
RA	Lucchini G., Lutzenkirchen K., Maat C.C., McConnell D., McKee R.A.,					
RA	Martegani E., Mathieu A., Maurer C.T.C., Montague M.A., Navas L.,					
RA	Messenguy F., Mewes H.-W., Mollema F., Montague M.A., Navas L.,					
RA	Newton C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,					
RA	Perea J., Philippsen P., Pierard A., Planta R.J., Plevani P.,					
RA	Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,					
RA	Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,					
RA	Sanz E., Schaaf-Gerstenschlaeger I., Scherens B., Schweitzer B.,					
RA	Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,					
RA	Spiegelberg R., Stave L.I., Steensma H.Y., Steiner S., Thierry A.,					
RA	Thieoes G., Triano L.N., Urrestazu L.A., Valle G., Vetter I.,					
RA	van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warrington J.R.,					
RA	von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,					
RA	Zimmermann F.K., Sgouros J.G.;					
RT	"The complete DNA sequence of yeast chromosome III.";					


```
FT VARSPLIC 4757 4761 TAHKS -> SKRRK (in isoform 2).
FT /FTID=VSP 003926.
FT VARSPLIC 4762 5147 Missing (in isoform 2).
FT /FTID=VSP 003927.
SQ SEQUENCE 5147 AA; 563537 MM; CD5D84990498CD3C CRC64;

Query Match 3.1%; Score 410.5; DB 1; Length 5147;
Best Local Similarity 19.0%; Pred. No. 5.1e-06;
Matches 477; Conservative 276; Mismatches 895; Indels 863; Gaps 105;

QY 552 DTSGEDNDEKAVASKGRKTANSQRRKGRITRSMANEANSEEAITPQOSAELASMELE 611
DB 8 DLSQSEERQIAA---VMGAAQGLPKGVSPPAAA-----ESPMSRKHQELSSHPK 58
QY 612 SSRWTEEMETAKGLLEH-----GRNWSAIARMVSGKTVSQCKNFYNYKK 658
DB 59 QSGRPDPGRPAQGLSKSRITDTFRSEQLPGRSPSTI-----SLKE 101
QY 659 RQNLDEILQOHLKME-----KERNARRKKKAPAAAASEEAPPPVVEEMEASGVGNE 714
DB 102 SKSRTDLKEEHKSMMPGFLSEVNA-----LSAVSSVNVKFNFPDLISDSEASQEETK 155
QY 715 EEWEEAEALHAG-----NEVPRGECSPATVNNSSDRESIPSPHTEAAKDTG 763
DB 156 KOKVVQKEQKPEGLIKPLOQQPKPIPKQGGPRDPLQDDGTPKSISSQOPEKIKS-- 213
QY 764 QNGKPPATLGADGPPGPPPTPRRTSRA-----PIEPTP 798
DB 214 ----QPGT---GKPIQGTQTPTQTHAKLPLQDASRPOTKQADIVRGESVPSLSP 265
QY 799 ASEATGAPTPPPAPSPAPPVVPKBEKETATAAPPVEGEEQKPPAAABELAVDTGKA 858
DB 266 SK-----PPIQPTPKPAQOGHEKSGPGPAKPPAQPSGLTKP-----LAQPGCTV 313
QY 859 EEPVKSECTEAEAE---GPAK-----GKDEAAEATAEGALKAEKKEGGSGRATTAKSS 909
DB 314 KPVVOPPTTKPPAQLPGLPAPPAQOTGSEKPSSE---OPGPKALAAQPPGVGK-TPAQOP 369
QY 910 GAPQDSSSATCSADEVDEAEGGDKNRLSPRLSLTPTGDPNANASPKPLDLKOLKOR 969
DB 370 G-----PAKPTQOVGTPKPLAQOGLQS 393
QY 970 AAATP-PIQVTKVHEPPREDAAATKAP-----PAPPPONTQPESDAP---QOPGSS-PRG 1021
DB 394 PAKAPGPTKTAQTKPPSQQSGSTKPPQPGPAKPSPO---QPGSTKPPSQOPGSAKPSA 451
QY 1022 KSRSPAPADKEAFAAABAKLPGDPPCWTSGLPVPVPREVIKASPHAPDPSAFSAPPG 1081
DB 452 QOPSPAPKS-----AQOF--TKPVSTQTFGKPLQPTV---SPSAKQ-----PPS 491
QY 1082 HPLP---LGLHDTARVLPRLPTIS-----NPPPLISSAKHPSVLBRQ 1121
DB 492 QGLPKTICPLNTWTELLHVPEKANFNCTECQITVCSLCGFNPNPHLTHEAKWMLNCQ 551
QY 1122 IGAISSQGMVOLVHPYSEHAKAPGVPT-----MGLPLP-----D 1157
DB 552 MKRALGGDLAP--VPSSPPKPLKTAPTVTTTSAVSKSSPOQOTSPKKDAAPKQDLSKAP 609
QY 1158 PKKLAPSGVQKQBL--SPRQOA-----GPPESLGVPQAQAS----- 1193
DB 610 PKKPPPL--VKOPTLHGSPSAKAKQPPREADSLSPAPPKPFSPVSEODKAFVADDKPKQP 667
QY 1194 --VLRGTALGVPGSITKGIPTSTKVP-----DSAITYRGSITHG---TPAD 1236
DB 668 KMWPTDULVSSSATTKPDIIPSSKVQQAQAEKTPPLKTDASKPSQFPPTGEKVTPTD 727
QY 1237 ---VLYKGTITRIICEPSRLDRGREDLSLPGHVIYEGKGHVLSEBGMSTVQCSKED 1293
DB 728 SKAIPRPASDLSIISHPGPSSSEKQKQVDP-----VQKKEPKKAQTKMSKPDAPKM 781
QY 1294 GRSSGGPHETAAKRTYDMMGRVGRRAISASIEGLMGRAPPERHSPHLLKQHHIR- 1352
DB 782 PKGSPTTP---GPRPT-----AGQTVPTPQOSP---KPQESRR 814

1353 QY ----GSITGIPRSYVEAOED-----YLRBEAKLLKREGTP----- 1384
815 DB FSLNLGSLTDA--PKSQPTTPOETVTGKLFGRGASIFSQASNLISTAGQPGHPSQSGPGAP 873
1395 QY -----PPPPSRDUTEAYKTOALGPLKUKPAHEGLVATVTKAG 1422
874 DB MKQAPASQPPTSQGPKPKSTQAPPAKASIPVKETKAPAAEKLEPAE--QAPTVKRTE 932
1423 QY RSIHEIP---REELRHTPELPLAPRLKEGSIQTQ--PL-----KYDTGASTTGSK 1469
933 DB TEKPPPPIKDKSKLTAEPQKAVLPKLEKSPKPESTCPLCKTELNIGSKDPNFTNCTEC 992
1470 QY KHDVRSLLIGSGRTFPFVPHPL-----DVMADARALERACYSLSKSRPGTASSS 1518
993 DB KNQVCNLG-----FNPTPHLTENCOTQRAISGQLDIRKMPPA--PSGPKASPMVPVTE 1045
1519 QY GGSARGAPVTVPELGRPRQSPPLTVEDHGAPFAGHLPRGSPVTMREPTP-----RLQ 1571
1046 DB SSSQKTAVPVQVKLVKQEQEVKTEAEKVI-----LEKVKETLSMEKIPPMVTTDQKQBE 1100
1572 QY GSLSSSKAS-----QDRKLTSTPREIAKSPHSTVP 1601
1101 DB SKLEKDKASALQEKKPPPEEKKLPIEEKIRSEKKPLLEKKPTPEKKLLPEAKTSAP 1160
1602 QY EHHPHPIPSYEHLLRGVSGVDLYRSHIPLAFDPTSIPIRGIPLDAAAAYVLPRLHAPNPTY 1661
1161 DB BEQKH-----DLKSKOVQIAEK-----LEGRVAPKTQV 1189
1662 QY PHLYPPVLIIRYGP--TAALENROTIINDYITSQOMHNTATAMAQRAADMRLGLSPRESS 1719
1190 DB EGKQPOTKMEGLPGSTPOSLEPKEDDKTTIKQEQPPCTAKPDQEKED-----DKSDT 1243
1720 QY LALNVAAGPRGIIDL-----SQVPHLPVLP-----PTGTP-- 1751
1244 DB SSSQOPKSPQLSGDTGYSSDGISSLSGLGIPSLIPTDEKDIILKGLKKDSFSQESSSPSPD 1303
1752 QY -----ATAMDLAYLPTAPQPPS-----SRHSSSPLSPGPGTHLTUKPTT 1790
1304 DB LAKLESTVLSTLEAQASTLADEKSEKKTQFHEVPEQPKDQEKTSQLEITLITISEEI 1363
1791 QY TSSSERERD-----RDRERKSILTSTTT-----VEHAFIWRPG 1830
1364 DB KESQERKDTFKKDSQDIPSSKHKEKSEFVDDITTRREPYDVSVEESSESSENSVPQRK 1423
1831 QY TEQSSGSGS-----SGGGG-----SSSRPASHSH-----AHOHSPISPR 1866
1424 DB RRTSVGSSSDEYKOEDSQSGSEBEDFIRKQIEMSDEADSGSDEDFIRNQLKEISS 1483
1867 QY TODALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSSPVVPAATFPFATHCPLGGTLD 1926
1484 DB TESQKE-----ETKKGKITAGKHRLT--RKSSTS----- 1513
1927 QY GVYPTLMEVLLPKEAPRVARPERPRADTGHAFIAPKPPARSGLBPASSPSKSGSEPRPLVP 1986
1514 DB -----IDEDA--GRHSHWDEDDAFDESP-----ELKYRETKSQSESEELV- 1552
1987 QY PVSGHATIAITPAKNLAPHASPDPPAPPASAPHREKTSQKPSI-----QELLES 2041
1553 DB -VTGGGLURFKTTEL-----NSTADKYSABSSQKTSLSYDFDEBEPELEMESL 1599
2042 QY -----GYHGSSYSPGVEFVPSVSPSLTHDKGLPKHLELDSKSHLEGSLR 2087
1600 DB TDSPEDSRSGEGSSSLHASSFTP---GTSPTSVSLLDED-----SDSPSHKKGSK 1648
2088 QY PKQPGVKLGEEAAHLPHRLPLPSQPS-----SPLQTAGVKGHQVRVVTQAHI 2139
1649 DB QOR-----KARHPHGPLLPITIEDSSEBEELREBEELKE---QEKREIBQQQRK 1696
2140 QY S-----EVITQDTRHHHPQOLSAPLPAPLPSFPGACSPVLDLRRPP--SDLYLPPP 2188
1697 DB SSSKKSKDKDELQARRRRRPK---TP-PSNLSPIEDAS--PTELROAEMEELHRSSC 1751
```


RESULT 26	PCLO_CHICK	STANDARD;	PRT; 5120 AA.
AC	Q9PU36;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Piccolo protein (Aczonin) (Fragment).		
OS	PCLO OR ACZ.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=99439764; PubMed=10508862;		
RA	Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,		
RA	Kilimann M.W.		
RT	Aczonin, a 550-kd putative scaffolding protein of presynaptic active		
RT	zones, shares homology regions with rim and bassoon and binds		
RT	profilin.";		
RL	J. Cell Biol. 147:151-162(1999).		
CC	!- FUNCTION: May act as a scaffolding protein involved in the		
CC	organization of synaptic active zones and in synaptic vesicle		
CC	trafficking (By similarity).		
CC	!- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).		
CC	!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of		
CC	synaptic junctions (By similarity).		
CC	!- DOMAIN: C2 domain 1 is involved in binding calcium and		
CC	phospholipids. Calcium binds with low affinity but with high		
CC	specificity and induces a large conformational change.		
CC	!- SIMILARITY: Contains 2 C2 domains.		
CC	!- SIMILARITY: Contains 1 PDZ/DHR domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; Y19187; CAB60725.1; -		
DR	HSP; P0410; 1A25.		
DR	GO; GO:0045202; C:synaptic junction; ISS.		
DR	GO; GO:0005509; F:calcium ion binding; ISS.		
DR	GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.		
DR	GO; GO:0005522; F:profilin binding; ISS.		
DR	GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.		
DR	GO; GO:0016080; P:synaptic vesicle targeting; ISS.		
DR	InterPro; IPR000008; C2.		
DR	InterPro; IPR001478; PDZ.		
DR	InterPro; IPR001565; Synaptotagmin.		
DR	InterPro; IPR008899; Znf_Piccolo.		
DR	Pfam; PF00168; C2; 2.		
DR	Pfam; PF00595; PDZ; 1.		
DR	Pfam; PF05715; Zf_piccolo; 2.		
DR	PRINTS; PR00399; SYNAPTOTAGMN.		
DR	SMART; SM00219; C2; 2.		
DR	SMART; SM00228; PDZ; 1.		
DR	PROSITE; PS00499; C2_DOMAIN_1; 1.		
DR	PROSITE; PS00004; C2_DOMAIN_2; 2.		
DR	PROSITE; PS01016; PDZ; 1.		
KW	Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;		
KW	Repeat.		
FT	NON TER	1	
FT	DOMAIN	258	357
FT	10 X 10 AA TANDEM APPROXIMATE REPEATS OF		
FT	P-A-K-P-Q-P-Q-Q-P-X.		
FT	ZN_FING	368	392
FT	ZN_FING	836	859
FT	DOMAIN	2324	2343
FT	POLY-PRO.		

Qy	881	--EAEATAEGALKAEBKKEGSGRAITAKSG--AFQDSSSATCS-----	922
Db	811	KHOTAEAKPAELSEQEKASQPKVCSPCLCKTGLNIGSKDPNFNFTTCECKKVCWNLGCFN	870
Qy	923	----ADBEDE-----AEGDKNRLSPRPSLLTPTGDPANASPOK-PUDLK	964
Db	871	PMFHIVEQWCLNCOTQRAMSQLGDMKVPLFKLPSQPIVSKP--PATPOKQPVAV	928
Qy	965	QLKQRAAAIPIQVTKVHEPPREDAAPTAPPAPPPQNLOPESDAPQPGSSPGKSR	1024
Db	929	SHSPQKSTFP---TPAATKPEEPSVPKEVPKLQCGKLEKTLISADKIQGIQKEDAKSK	985
Qy	1025	S-----PAPPAD-----KEAFABAOKLPEDPPC-----WTSGLPFPVPPR	1060
Db	986	QGKLFKTPSADKIORVSKEDSRLQQTCLKTFTSPSSDKILHGVOKEDIKFOEAKLAKTPSA	1045
Qy	1061	EVIKASPHADP-----SAFSYAPGHPPLPLGLHDT-----	1091
Db	1046	DKILHRLQKEDPKLQOMMAKALSADKIQPEAKEDVQLQEVRLSKAVSADKIQHGIQKD	1105
Qy	1092	-----ARPVLR-----PPTISN-----	1104
Db	1106	LNLOHVKIEXTSSVEKIQEAKESKLOQDKLPKTLSEDKIPATVSSDHHKLLSKSEBDKK	1165
Qy	1105	PPPLISAKHPVLEROIGAISQ-----MSVQLHVPYSE-HAK-----APGVPVTMGLP-L	1154
Db	1166	PELLEKSTPHPKDKEQITAEITTHIPEQKVEVEAPCDLHEKQEDVKKEDLTTGIPQM	1225
Qy	1155	PMPOKPLAPSGVKQEQOLSPRGQAGPESLGVPTAOE-----ASVLRGTA LCS	1202
Db	1226	VSKPEK-----ABEEKTPVPSRLPRSDHVEAREKIEKEDKSDTSSQQKSPQGL	1278
Qy	1203	VPGGSITKGIPS-----TRVPDSAITYRGSITHGTPADVLVKGITRIIGEDSPSRLD	1256
Db	1279	SDTGYSDGSISSIGEIPSHIPEDE-----KOLPRSPSQDTISO---ESPSPSD	1326
Qy	1257	RGREDG-----LPGKH-----	1267
Db	1327	LAKLESTLILEAQASTLDEKSVKRKELYETYSEQTKDQHKTKPLPVTPEVSYSDEED	1386
Qy	1268	--VIEYCKGHVLSYEGG-MSVTCQCKEDGRSSGPPHETAAPKRTYDMMEGRVGRAISS	1324
Db	1387	LEAIQEGERTIAADSKGASSQTDYKEEDG-----GNDTPARRQRYSVED--SSESEN	1438
Qy	1325	ASTEGLMGRAIPIPERHSPHLKEQHHRIGSITOGIPRSYVE--AOED-----YLRR	1374
Db	1439	SPVPRKRRASVSSSSDEYKRDSDQSGDEEDFIRKQIITEMADEDASGEDEDFIRNQ	1498
Qy	1375	AKLLKREGTTPPPPSRDLTEAYKTOALGPLKPAHEGLVATVK-----EAG	1422
Db	1499	LKEI-----SVTESQKE-----EVKSKAGTVGKRRMARKSSAGYDEDAG	1540
Qy	1423	R--SIHEIPREELRHTPELPIAPRLPKEGSITOGTPLKYDTGASTTCKGKHVRSLLGSP	1480
Db	1541	RRHSWHDDEDTDESPE----PK-YRETKSQDGEELAISGGGGLRRFKTIELNSTITSK	1595
Qy	1481	GRTFPPVHPLDMADARALERACRYEESLSKRPCTGASSGSGSIARGAPVIVPELGKPRQS	1540
Db	1596	YSPTPEQQKILYFDE--EPELEMESLTOSPEDRSRGEGSSSIHASSFTPTGTSPTSVSS	1652
Qy	1541	LTYPEDHGAPF-----AGHLPRG-----SPVTMREPTPRLOE-----	1571
Db	1653	LDESDSSPSHKKLGGESKQQRKARHRSHGFLPTIEDSFEELREEBELLKXEQEKORE	1712
Qy	1572	-----GSLUSSKASODRKLTSTPREIAKSPHSTVPEHHPHPI--SPVEHLLRGVSGV	1621
Db	1713	LEQQRKSSSKSKKODLELRAORRRRPRKTPPSNL-----SPTEASPTFEELRQAEME	1767
Qy	1622	DLYRSHIPLAFDPT--SIRGIP-----LDAAAAAYLPRHLA-PMPT-----	1660
Db	1768	ELHRSSCS--EYSPSIESDPGEFELSPEKILVOKVVKLPFAVLSYSPSTDEKLTGALKES	1826
Qy	1661	-----YPHLYPPYLIRGYDPTAALEN---1681	

RESULT 27

CC This SWISS-PROT entry is copyright It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; U52950; AB17068.1; --
CC EMBL; X60370; CAC16162.1; --
CC EMBL; X16623; CRA34620.1; ALT SEQ.
CC PIR; A56577; A56577.
CC InterPro; IPR00102; MAP1B neuraxin.
CC Pfam; PF00414; MAP1B neuraxin; 10.
CC PROSITE; PS00230; MAP1B_NEURAXIN; 8.
CC Microtubule; Repeat; Phosphorylation.
CC CHAIN ? 2459 MAP1 LIGHT CHAIN LC1.
CC FT REPEAT 1869 1885 MAP1B 1.
CC FT REPEAT 1886 1902 MAP1B 2.
CC FT REPEAT 1903 1919 MAP1B 3.
CC FT REPEAT 1920 1936 MAP1B 4.
CC FT REPEAT 1937 1953 MAP1B 5.
CC FT REPEAT 1954 1970 MAP1B 6.
CC FT REPEAT 1988 2004 MAP1B 7.
CC FT REPEAT 2005 2021 MAP1B 8.
CC FT REPEAT 2022 2038 MAP1B 9.
CC FT REPEAT 2039 2055 MAP1B 10.
CC FT DOMAIN 559 1035 GLU-RICH.
CC FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
CC KKEE AND KKEI/V REPEATS).
CC FT DOMAIN 1224 2312 LYS-RICH.
CC FT CONFLICT 127 127 M -> V (IN REF. 1).
CC FT CONFLICT 140 140 T -> S (IN REF. 1).
CC FT CONFLICT 2112 2112 R -> K (IN REF. 3).
CC FT CONFLICT 2169 2169 L -> I (IN REF. 3).
CC FT SEQUENCE 2459 AA; 269497 MW; 2E3F6872DEDB8BA2 CRC64;
CC SQ
CC
CC Query Match 3.1%; Score 406.5; DB 1; Length 2459;
CC Best Local Similarity 19.3%; Pred. No. 3.4e-06;
CC Matches 489; Conservative 326; Mismatches 995; Indels 729; Gaps 107;
CC
CC QY 68 LLSEFQGNERSOEL-----HLRPESHLYLPEL-GKSEM----- 100
CC DB 174 LLSTTHPANKASLTFCPEEGDKNSLDRLHDLQFINIKLSASILPEMEGLSEFTYL 233
CC QY 101 -EFIESKRPRLELLDP-----LLRP-----SPLATGQ-----PAGEED-- 134
CC DB 234 SESVEVPSP-FDILEPPTSGGFLUKSKPCCYIPGCGDSALFVNGFNMLINGSERKS 292
CC QY 135 -----LTK--DRSLTG-----KLEPVSPSPPHTD-----PE 159
CC DB 293 CFWKLRLHLDVDSILLTHIGDNLPGINSMLQRLKIAELSESGSTNSDMKNL19PD 352
CC QY 160 LEL-----VPRRLKEELIQMDRVREITWVEQOISLKKKQOOLEEAAKPEP----- 211
CC DB 353 LGVFLVNPENLKNPEFNKKESTBEACFTQYLNKLSNKKPEPLFRSNGVNAIEPVILFQ 412
CC QY 212 K-----PVSPPTIESKHSRLVQIIVDENRKAFAAHRILEGLGPOVELPLYNQPSD 262
CC DB 413 KMGVGLKKNVLPNPKSSKEMQYFMQQTGTNKKAE-----LILPNGQEVDP----- 461
CC QY 263 TRQYHENIKINQAMRKKLILYFKRRNRHARKQWKQKFCQYDQ--LMEALEKKVERLENNP 320
CC DB 462 -----ISYLASVSLIVWHPANPAEKIIRVLPFGNSTQNIILEGL-LKHLDFLK 511
CC QY 321 RRRAKESKVEYVEKQFPEIRKQRELOERQMSRGVORGSGLSMSAARSEHSEVSEIIDGUS 380
CC DB 512 QPLATQKDLTG--QVSTPPV-KQVCLKQKADRESLKLKATPKLSSKSVKESKEAPEAT 568
CC QY 381 EQENLEKQMRQLAVIPMLVDYDQORIKFTINMGLMADPNKYVKDR--QVMNMSQKE 438
CC DB 569 KASQVEK-----TPKVESKEKVIYK-----KDKPKVESKPSVTEKE 605
CC
CC MAP1B RAT
CC ID MAPR RAT STANDARD; PRT; 2459 AA.
CC AC P15205; Q62958; Q9ER21; Q9QW92;
CC 01-APR-1990 (Rel. 14, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
CC DE light chain LC1].
CC GN MAP1B.
CC OS Rattus norvegicus (Rat).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC OX NCBI_TaxID=10116;
CC [1]
CC RP SEQUENCE OF 1-142 FROM N.A.
CC RC STRAIN-Sprague-Dawley; TISSUE=Testis;
CC MEDLINE=96357242; PubMed=8666295;
CC RA Liu D., Fischer I.;
CC RT "Isolation and sequencing of the 5' end of the rat microtubule-
CC associated protein (MAP1B)-encoding cDNA.";
CC Gene 172:307-308(1996).
CC RL [2]
CC RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
CC RC STRAIN-Sprague-Dawley; TISSUE=Brain, and Glial tumor;
CC MEDLINE=92347374; PubMed=1639092;
CC RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
CC RT "Identification of two distinct microtubule binding domains on
CC recombinant rat MAP 1B.";
CC Eur. J. Cell Biol. 57:66-74(1992).
CC RL [3]
CC RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
CC RC TISSUE=Spinal cord;
CC MEDLINE=90059871; PubMed=2555150;
CC RA Rientz A., Gremmling G., Hermans-Borgmeyer I., Kirsch J.,
CC RA Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
CC RT "Neuraxin, a novel putative structural protein of the rat central
CC nervous system that is immunologically related to microtubule-
CC associated protein 5.";
CC EMBO J. 8:2879-2888(1989).
CC RL [4]
CC RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
CC MA D., Nothias F., Boyne L.J., Fischer I.;
CC RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
CC in rat CNS and PNS during development.";
CC J. Neurosci. Res. 49:319-332(1997).
CC CC -!- FUNCTION: The function of brain MAPS is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC CC -!- INDUCTION: By nerve growth factor.
CC CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (by similarity).
CC CC -!- PTM: Phosphorylated.
CC CC -!- SIMILARITY: TO MAP1A.
CC CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC 2459) was originally described as neuraxin in Ref.3.

Mon Apr 19 08:15:07 2004

DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RL Biochem. J. 253:919-922(1988)."
RN [2]
RN SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA Myers J., Williams C., Ramirez F.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RL conservation of a pattern of introns and exons.";
RN Nature 310:337-340(1984)."
RN [3]
RN SEQUENCE OF 162-301.
RC TISSUE=Skin;
RX MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RL the alpha 1 and alpha 2 chains of human skin collagen.";
RN Biochemistry 9:4699-4706(1970)."
RN [4]
RN SEQUENCE OF 263-268.
RC TISSUE=Skin;
RX MEDLINE=71001508; PubMed=4319110;
RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT "A comparative study of glycopeptides derived from selected
RL vertebrate collagens. A possible role of the carbohydrate in fibril
RN formation.";
RN J. Biol. Chem. 245:5042-5048(1970)."
RN [5]
RN SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;
RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA Prockop D.J.;
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
RL pro alpha 1 chain of human type I procollagen. Statistical evaluation
RN of structures that are conserved during evolution.";
RN Biochemistry 22:5213-5223(1983)."
RN [6]
RN SEQUENCE OF 1229-1454 FROM N.A.
RC TISSUE=Bone;
RX MEDLINE=88124208; PubMed=3340531;
RA Mäkelä J.K., Raassina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RL domain.";
RN Nucleic Acids Res. 16:349-349(1988)."
RN [7]
RN SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinak R.E.;
RT "Regulatory elements in the first intron contribute to
RL transcriptional control of the human alpha 1(I) collagen gene.";
RN Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987)."
RN [8]
RN SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RL Promoter structure, Alu repeats, and polymorphic transcripts.";
RN J. Biol. Chem. 260:2315-2320(1985)."
RN [9]
RN SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.J.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RL collagen gene enhance transcription.";
RN J. Biol. Chem. 262:15151-15157(1987)."
RN [10]
RN REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RL in humans.";
RN FASEB J. 5:2052-2060(1991)."
RN [11]
RN REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RL associated collagen (type IX), and network-forming collagen (type X)
RN cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RN Hum. Mutat. 9:300-315(1997)."
RN [12]
RN REVIEW ON VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RA Byers P.H., Wallis G.A., Willing M.C.;
RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
RN J. Med. Genet. 28:433-442(1991)."
RN [13]
RN REVIEW ON VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RA Dalglish R.;
RT "The human type I collagen mutation database.";
RN Nucleic Acids Res. 25:181-187(1997)."
RN [14]
RN VARIANT OI-II CYS-1166.
RX MEDLINE=86287390; PubMed=3016737;
RA Cohn D.H., Byers P.H., Steinmann B., Gelinak R.E.;
RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
RL change in one human pro alpha 1(I) collagen allele.";
RN Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986)."
RN [15]
RN VARIANT OI-II ARG-569.
RX MEDLINE=87222295; PubMed=3108247;
RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
RL arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RN collagen.";
RN J. Biol. Chem. 262:7021-7027(1987)."
RN [16]
RN VARIANT OI-II CYS-926.
RX MEDLINE=88033031; PubMed=3667599;
RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT "A point mutation in a type I procollagen gene converts glycine 748
RL of the alpha 1 chain to cysteine and destabilizes the triple helix in
RN a lethal variant of osteogenesis imperfecta.";
RN J. Biol. Chem. 262:14737-14744(1987)."
RN [17]
RN VARIANT OI-II ARG-842.
RX MEDLINE=88298828; PubMed=3403550;
RA Bateman J.F., Lemande S.R., Dahl H.H., Chan D., Cole W.G.;
RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
RL chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RN the peptide defect by in vitro expression of the mutant cDNA.";
RN J. Biol. Chem. 263:11627-11630(1988)."
RN [18]
RN VARIANT OI CYS-1195.
RX MEDLINE=89218628; PubMed=3244312;
RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT "A cysteine for glycine substitution at position 1017 in an alpha
RL 1(I) chain of type I collagen in a patient with mild dominantly
RN inherited osteogenesis imperfecta.";

RL	Mol. Biol. Med. 5:197-207(1988).	Qy	918	SATCSADEVDEAEGDKNRLSP----	RPSLLTTPGDPRANASPKQLDLKQLKORAAAI	973
RP	[19]	Db	283	-----AGPKGEGPCGENGAPGQMGPRGLPGERGRPGAP-----	GPAGAR	322
RA	VARIANT OI-II VAL-434.	Qy	974	PPIQVTKVHEPPREDAAPTKA-PPAPPPQNLQPEDAPQPGSS--	PRGKSRSAPP-	1029
RA	MEDLINE=89255493; PubMed=2470760;	Db	323	GNDGATGAAGPP-----GPTGAGPPGPGFAGVAKGEA-GPQGRGSEGPQVRGEPGPPG	377	
RT	"RNA sequence analysis of a perinatal lethal osteogenesis imperfecta mutation.";	Qy	1030	-----ADKEAFAAEAQKLPDPPCWTGLP-FVPVPREVIKASHAPDPSAFS	1076	
RL	J. Biol. Chem. 264:10083-10087(1989).	Db	378	PAGAAGPAGNPGADGQPCAKCANGAP-----	IAGAPFPARGSPGPGGPPGPKGN	432
RP	VARIANT OI-IV SER-1010.	Qy	1077	YAPGHPPLPLGLHTARPVLPRPTISNPPPLISSAKHPSVLERQIGAISQMSVOLHVP	1136	
RA	MEDLINE=89308591; PubMed=2745420;	Db	433	SGEPGAPGSKG-DTGAKGERGPGVVOGPP-----	GPAGEGRG	470
RA	Marini J.C., Grange D.K., Cottresman G.S., Lewis M.B., Koepf D.A.;	Qy	1137	YSEHAKAPVGPVTMGLPLPMDPKLAPSGVQKQLSPRGQAGPESLGVTAQEAASVLR	1196	
RA	"Osteogenesis imperfecta type IV. Detection of a point mutation in one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";	Db	471	-----ARGEPGT--GLP-----	-----GPPGERGGPGSRGFPDAGVAGPK	505
RL	J. Biol. Chem. 264:11893-11900(1989).	Qy	1197	GTA--LGSVPGSITKGP-STRVPSDAITYRGSITHTPADVLYKGTITRIIGEDSPS	1253	
RL	J. Biol. Chem. 264:15809-15812(1989).	Db	506	GPAGERS-PGAPGPKSPGEAGRPGEAGLPCAAGLT-GSPGSPGPDGK----	TGPPGPA	559
RP	VARIANT OI SER-1022.	Qy	1254	RLD-----RGREDSL-----PKGHVIEGKKGHLVSYEGM-----	SVTQCSKED	1293
RA	MEDLINE=90062068; PubMed=2511192;	Db	560	GODGRPGPPGPGARGQAGVGMFPKGAAGEPGKAG-----	ERGVPPPGAVGPAKG	614
RT	"Substitution of serine for alpha 1(I)-glycine 844 in a severe variant of osteogenesis imperfecta minimally destabilizes the triple helix of type I procollagen. The effects of glycine substitutions on thermal stability are either position of amino acid specific.";	Qy	1294	GRSSGGPHETAARKTYDMGHRVGRASISASIEGLMGRAPP-ERHSPHHLKEQHHR	1352	
RL	J. Biol. Chem. 264:19694-19699(1989).	Db	615	EAGAQGP-GPAGPAGE-----RGEQGA-GSPGQGLPGPAGPCEAGKPE-----	660	
RP	VARIANT OI-II CYS-1082.	Qy	1353	GSITQIGIPRSYVEAQEDYLREALKKREGTPPPPSRDITEAVKTQALGPLK----	PA	1410
RA	MEDLINE=89109573; PubMed=2913053;	Db	661	-----QGVPLDLCAPGSPGARGERFPGERGVOGPPGA-----	GRGANGAPG	704
RA	Constantinou C.D., Nielsen K.B., Prockop D.J.;	Qy	1411	HEGLVATVKEAGRSIHEIPREELRHTPELPLAPRLKEGSITQSTP-----	LKYDTGASTT	1466
RT	"A lethal variant of osteogenesis imperfecta has a single base mutation that substitutes cysteine for glycine 904 of the alpha 1(I) chain of type I procollagen. The asymptomatic mother has an unidentified mutation producing an overmodified and unstable type I procollagen.";	Db	705	NDGAKG--DAG-----APCAP-----GS--QCAPGLQGMPPGERGAAGL	738	
RL	J. Clin. Invest. 83:574-584(1989).	Qy	1467	GSKXHD-----VRSLSGSGR-----TFP--PVHPLDVMADARALERACVEESLKSPPG	1513	
RP	VARIANTS OI CYS-272; CYS-704 AND CYS-896.	Db	739	PGPKGDRGDAGPKGADGSGPKDGVRLTGTPGPPGAPAGD-----	KGESG	785
RA	MEDLINE=90009313; PubMed=2794057;	Qy	1514	TASSSGSISARGAVIPVELGPKRQSLTYEDHCAFPAGHLPRGSPVTMREPTPRELQGS	1573	
RA	Stearns B.J., Eyre D., Charbonneau H., Harrylock M., Weiss M.A., Weiss L., Graham J.M., Byers P.H.;	Db	786	PSGAPGTGARGAPGDRGEGPP--GPAGFA--GPPGAD-----GQPKAGBPGDAGAKGD	837	
RT	"Osteogenesis imperfecta. The position of substitution for glycine by cysteine in the triple helical domain of the pro alpha 1(I) chains of type I collagen determines the clinical phenotype.";	Qy	1574	LSSSKASQDRKLTSTPREIAKSPHSTVPEHHHPPI-----SPYEHLRLGVSGVDLYRSHIP	1629	
RL	J. Clin. Invest. 84:1206-1214(1989).	Db	838	-----AGPPGAPGAPGPPGPGI GNVGAPGAKGARGSAGP-----	870	
RP	VARIANT OI-II CYS-422.	Qy	1630	LAFDPTSIPIRGIPLDAAAAYLPHLAPNPTYPHYLIRGYPTDPALENROTII INDY	1689	
RP	Query Match 3.0%; Score 400.5; DB 1; Length 1464;	Db	871	-----PGATGPPGAGRVGP-----PGPS-GNAGPP-----GPPGAPGEGGKGRGE-	912	
RP	Best Local Similarity 21.5%; Pred. NO. 3.3e-06;	Qy	1690	ITSQMHNTATAMAQADML-----RGLSPRESSALNYAAGPGIIDLQVPP	1738	
RP	Matches 380; Conservative 102; Mismatches 642; Indels 647; Gaps 83;	Db	913	-----TGAPRPGVEGPPGPPGAPGKSGPADGAPAGPTGPGIAGQGVV	961	
Qy	722 BALHASNEVPRGCSGATVNNSSDTSISPHTEAAK-DTGQNGKPPA-TIGADGPP	Qy	1739	HLPVLPVPT--PGTATAMDLAYLPTAPQFSSRSHSSPLSPGSPHLLTKPTTTSSSR	1796	
Db	77 ETKNCPGAEPGEGCC-PVCPDGSESPDQTTGVEGPGKGTGPRGRGAPGPPGRDIP	Db	962	GLPQQRGERGPPGLPGPSGE-----PGKQPGSASGERGPPGPPGCLAGPPGESGRE-	1015	
Qy	.780 -----PGPPTTPRTRAPIEFTPASEATGAPTPP-----	Qy	1797	ERDRDRDRDREREKSLITTTTVEHAPIWRPQTEOSSSGSSSGSGGGG-----	GSSSRP	1851
Db	136 GQPLGPPGPP-----	Db	1016	-----GAPGAEGSPGRDGSFGAKGRGTGAPGP	1045	
Qy	811 APPSPSAP-----PPVVPKSEKETAAAPVVEGE-----	Qy	1852	ASHSHAHQHSPIPRITQDALQQRPSVLHNTGKGIITAVEPSKPTVLRSTSTSPVRPA	1911	
Db	179 GPMGPGSRGLPFPAGPQCFQ-----GPPGEPCEPGASGPMGPPGPPGKNGDDGE	Db	1046	GAPGAPGAPGVP-----	AGVPVGPAG	1082
Qy	858 ABEVPKSECTEAEAGPKADAEAEATAGALKAEKKEGSGRATTAKSSGAPQSDS	Qy	1912	TFPPATHCPLGGTLDGVYPTLMPEVLLPKCAPRVARPERPRADTGH-----	1957	
Db	235 AKPGR-----PGERGPPPGQARGLPCTA--GLPCKMKGHRGFGSLDGAKGDAGP-----	Db				

MLL3	MOUSE	
MLL3	STANDARD; PRT: 4903 AA.	
Q8BRH4	Q8BK12; Q8C6M3; Q923H5; Q923H6;	
AC	10-OCT-2003 (Rel. 42, Created)	
DT	10-OCT-2003 (Rel. 42, Last sequence update)	
DDT	15-MAR-2004 (Rel. 43, Last annotation update)	
DE	Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog (Histone-lysine N-methyltransferase, H3 lysine-4 specific MLL3) (EC 2.1.1.43).	
GN	MLL3.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
NCBI_TaxID=10090;		
[1]	CONCEPTUAL TRANSLATION OF 814-956 AND 1377-4213.	
RP	Blatter M.-C.;	
RP	Unpublished observations (SEP-2003).	
RL	[2]	
RP	SEQUENCE OF 957-1376 AND 4214-4894 FROM N.A.	
RC	TISSUE=Myeloma;	
RC	MEDLINE=21574953; PubMed=11718452;	
RC	Tan Y.C., Chow V.T.;	
RT	"Novel human HALLR (MLL3) gene encodes a protein homologous to ALR and to ALL-1 involved in leukemia, and maps to chromosome 7q36 associated with leukemia and developmental defects.";	
RT	Cancer Detect. Prev. 25:454-469(2001).	
RL	[3]	
RP	SEQUENCE OF 1-813 AND 4803-4903 FROM N.A.	
RP	STRAIN=C57BL/6J; TISSUE=Embryo;	
RC	MEDLINE=22354683; PubMed=12456851;	
RC	Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,	
RA	Nikaudo I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,	
RA	Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojibori T.,	
RA	Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,	
RA	Schriml L.M., Kanapin A., Matsuda H., Batalov D., Corbani L.E., Cousins S.,	
RA	Blake J.A., Bradt D., Brusic V., Fletcher C.F., Forrest A., Frazer K.S.,	
RA	Dalla E., Dragani T.A., Godzik A., Godzik A., Gough J.,	
RA	Caasterland T., Gariboldi M., Gissi C., Jackson I.J., Jarvis E.D.,	
RA	Grimmond S., Gustinich S., Hirokawa N., Kedzierski R.M., King B.L.,	
RA	Kanai A., Kawaji H., Kawasaki Y., Lee Y., Lenhard B., Lyons P.A.,	
RA	Konogaya A., Kurochkin I.V., Marchionni L., McKenzie L., Miki H.,	
RA	Magliott D.R., Maltais L., Pavan W.J., Pertea G., Pesole G.,	
RA	Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,	
RA	Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,	
RA	Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,	
RA	Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,	
RA	Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,	
RA	Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,	
RA	Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,	
RA	Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayateu N.,	
RA	Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,	
RA	Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,	
RA	Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,	
RA	Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,	
RA	Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,	
RA	Birney E., Hayashizaki Y.;	
RT	"Analysis of the mouse transcriptome based on functional annotation of	
RT	60,770 full-length cDNAs.";	
RL	Nature 420:563-573(2002).	
CC	-1- FUNCTION: Belongs to the ASC-2/NCOA6 complex (ASCOM), a	
CC	coactivator complex of nuclear receptors, involved in	
CC	transcriptional coactivation. MLL3 may be a catalytic subunit of	
CC	this complex, which weakly methylates Lys-4 of histone H3. This is	
CC	a specific tag for epigenetic transcriptional activation (By	
CC	similarity).	
CC	-1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =	
CC	S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.	
CC	-1- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which	
CC	contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/	
CC	RBBP5, alpha- and beta-tubulins, the trithorax group proteins MLL2	
CC	and MLL3, and ASH2/ASCL2. Interacts with histone H3 (By	
CC	similarity).	
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).	

Db	3291	TSLTMSQP--NFMVPMQOQHQHTAVISGHTSPARMPSLPGWQSNAS-----	3337	OS	Homo sapiens (Human).
Qy	1731	IIDLSQVPHLPVLP-----PTGCTATAMDRLAYLPTAPQPPSSRHSS 1775		OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Db	3338	-----AHLPLNPRTOPPIAQLSLKCTCTPACTVSSANPQ-----NGPPPRVFDDNN 3385		OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Qy	1776	PLSPGGPHLTPTTSSRER-----DRDRDRDRERK-----SILTSTTTVEH 1823		RN	NCBI_TaxID=9606;
Db	3386	PFSESQERERKERLREQERQVQLMQEVDQRALQORMEQHCLMGAEALANRPVSQ 3445		RP	SEQUENCE FROM N.A. (ISOFORM A).
Qy	1824	APIWRPTEQSGSGSGSGGGGSSRPA-----SHSHAHQHSPI-----		RC	TISSUE=Lung;
Db	3446	MPFY-----GSDRPFCDLPQPRPLOSPOHQOIGPVLOQOQVOQG 3486		RX	MEDLINE=96070776; PubMed=7592926;
Qy	1864	---SPRTQDALQ---OR---PSVLHN---TGKGIITAVEPSKPTVLRSTSTSSVPR 1909		RA	Miura Y., Tam T., Ido A., Morinaga T., Miki T., Hashimoto T.,
Db	3487	SVNSPPNQTQWTEQRQVGFPSFVDPSPSAGSPNFKVKGHNLPGSSFOQSPLRP 3546		RA	Tamaoki T.;
Qy	1910	-----AATFPPA-THCEL-----GGTLDGVVPTLME---PVLLPKAPRVARPERPR 1952		RT	"Cloning and characterization of an ATBF1 isoform that expresses in a
Db	3547	PFTPIPGTSPVANSNVPCGQDPVATQONYSGSSQSLIQLYSIDIPEEKGGKTRTKKK 3606		RT	neuronal differentiation-dependent manner.";
Qy	1953	ADTHAFIAPKPARSGLEPASPSKSEPRPLVPVSGHATIAITPAKNLAPHASDP- 2011		RL	J. Biol. Chem. 270:26840-26848 (1995).
Db	3607	KD-----DDAESGKAPSTPHSDCA--APLTGCLSETTTPAVSPSELPOQRQEPV 3656		RN	SEQUENCE FROM N.A. (ISOFORM B).
Qy	2012	-----PAPPASADPHREKTOSKPPS--IOELELRSLGYHSGSSYSPGCVPSVSPSL 2064		RC	TISSUE=Hepatoma;
Db	3657	EPVPVPTNVASGQCIENSEKLPNSFEIKETSNQOOTHVNAEADKPS-----VETPNK 3709		RX	MEDLINE=99425270; PubMed=10493829;
Qy	2065	THDKGLPKHLELDKSHLEGEURPKQGPVK--LGGEAAHLPLRLPLPESOPSSPL-LQ 2121		RA	Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
Db	3710	TEEIKLEK--AETQPSQEDTKVEETGNKIKDIVAG-----PVSSIQCPHPVGT 3758		RA	Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
Qy	2122	TAPGVKHQVVTLAQHI-----SEVITQDYTRHHQOOLSAPLAPLYSPFGASCPVLD 2175		RA	Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
Db	3759	TKGDTGNE-----LLKHLKNKASLLTQ-----		RA	Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
Qy	2176	LRPPSGLYPPDHGAPARGSPHSGGKRSEPNKTSVLGGGEG--IEPVSPPEGNTE 2233		RT	"Genome duplications and other features in 12 Mb of DNA sequence from
Db	3785	-----KPEGLTSSDESSTKDGKLEIKQSPAEGLQT 3814		RT	human chromosome 16p and 16q.";
Qy	2234	PHSRNAVY-----PLLYRDGEOTEPSPRMGSKSPGNTSOPPAFFSKLTESNAMYKSKQ 2289		RL	Genomics 60:295-308 (1999).
Db	3815	LGAQMCGGFGGNSLQPKTDG--ASENKKORSKRTQTEKAAPRSKRRKQ---EEBKQ 3869		RP	SEQUENCE OF 1151-3703 FROM N.A. (ISOFORM A).
Qy	2289	EINKKLNTHNRNPEYNISQGTETEFNMPAITGTGLMTRYSOAVQSHASTNMGLEAI--- 2345		RA	Kozlovicz A., McQuerry Y., Hotic M.;
Db	3870	AMYSSDSFTHLQONNLNPNPTPPASLPP-----TPPPMACQRMANGFATTEELAGK 3922		RA	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
Qy	2346	-----IRKALMGKYDQWEESE--PLSANAFNPLNASALPAAAMPITAAQDRSDHTLT 2395		CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
Db	3923	AGVLVSHEVARALGPKPQLPFPFQDILLARAIAQGPKTVDVPASLPTPHNNHELRQ 3982		CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
Qy	2396	SPGGGKAKVSGRRSPSKAKSPAPGLASGDRPPSVSVHSEGDGNCNRTPLTRVWEDRPS 2455		CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
Db	3983	DHYGDRTPDSFVPS-----SPESVGVGEV-----NKYPDLSLVKEEPPE 4023		CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
Qy	2456	SAGSTPPFYNPLI-----MRLQAGVMASPPPPGLPAGSGPLAG 2493		CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
Db	4024	PVPSPIIPILPISGKNSESRNDIKTEPGTLFTTSPFG--SSPGRSG 4071		CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RESULT 31				CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
ABF1 HUMAN				CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
ID	ABF1	HUMAN	STANDARD;	CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
AC	Q15911	Q15911; Q13719;	PRT; 3703 AA.	CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DT	16-OCT-2001	(Rel. 40, Created)		CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DT	16-OCT-2001	(Rel. 40, Last sequence update)		CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DT	10-OCT-2003	(Rel. 42, Last annotation update)		CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DE	Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)			CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DE	(AT-binding transcription factor 1).			CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
GN	ATBF1.			CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

QY	956	SPOKPLDL-KQLKORAAAIPTQVTKVHEP-----PREDAAPTKP-----APPAPB---PP	1002
Db	2386	MPSQAYSAPAPSANNATSSAFIQLTAEBELATFNSTAGDEKPLAEASQAPNQTOE	2445
QY	1003	QNLQPSDAPQOPGSGSPRGKSRSPAPPADKEAFAEAQKLPDGPCCWTGULFPVPPREV	1062
Db	2446	KQOQPRPELQOO--EQPEQKTNP-----QQKLP-----	2472
QY	1063	IKASPHAPDPSAFSYAPPGHPLPLGHDHTARPVLPRPPTISNPP--PLISSAKHPSVLER	1120
Db	2473	-----QLVSLPSLPQPPQAPPPQCPFLPQSGSPSPQSLS-2505	
QY	1121	QIGAISQGSVQLHLVYSE-HAKAPVGMTG--LPLPMDPKLA--PFGSVKOEQ---	1171
Db	2506	-----HLPLKPLHTSTPOOLANLPPOLIPLYOCDOCKLAFSPFEHWQSHQQLH	2552
QY	1172	LSRGQAGPPE---SLGVP-----TAQEASVLRGTALGSVPGSGITKGIPTSRVPS	1219
Db	2553	FLSAQNQFIHQFLDRSLDMFPLFDPSPNLLASQLLSGAIPQIPASSATS--PST--PT	2608
QY	1220	DSAITYEGSIHTGTPADVLKGTITRIIGDPSRLDRGREDSLPGK-HVIVEGKKGHVL	1278
Db	2609	STMWTLKRKLEEKASAS---PGENDSGTGEEEPQDRKLRTITPQLEILIQ---KYL	2662
QY	1279	SYEGGSMVTCQSKEDGRSSGPPHETAAPKRTYDM-----MEGRVGRAISSASIEGLMG	1332
Db	2663	D-----SNPTRKWLHDIAHEVGLKRVQVWVFQNTRARERKQF-----	2701
QY	1333	RAIDP---ERHSPH-----HLKEOH-----HJR	1352
Db	2702	RAVGAQAHRCPFCRALFKAKTALAEHTSRHWEAKRAGYNLTUSAMLLDCDGLQWK	2761
QY	1353	GSITQIGIPRSYVEAQEDYLREAKLLKRECTPPPPPSRD-----LTEAYKTQALGPLK	1407
Db	2762	GDI FDGTSFSL-----PPSSSDGQGVLPSPVSKTMELSPRTL	2799
QY	1408	KPAHEGLVATVKEAGRIHEIPREELRHTPELPLAPRPLKEGSIOTGTPKVDYTGASTTG	1467
Db	2800	LSP-----SSIKVEGIEDFESP---MSSVNLNFDQTKLNDNDCSSVNTAITDTTGDGSG	2851
QY	1468	SKKHDVRSLLIGPGKRTFPVHPLDVMADARALERACVYESLKRPGPTASSSGSARGAP	1527
Db	2852	NADND--SATGIATETKSSAPNEGLTKAAMWAMSEYEDRLSS--GLVSPA-----	2898
QY	1528	VIVPELGPKPROSPITYEDHGAPFAGHLPGRSPVTMREPTPRLQEGSLSSSKASQDR---	1583
Db	2899	-----PSFYSKEYNEGTV--VDYGETSLADPCSPSPG--ASGSAGKSGDGDPRGQK	2947
QY	1584	---KLSTPREIAKSPHSTVPEHHHPHPISPYEHLRGSVDLYRSHIPLAFDPTSI	1639
Db	2948	RFTQMTNLQKVLKSCFND-----YRPTMLECEVLGNDI	2983
QY	1640	GIPLDAAAAY-----LPRHLAPNPT-YPHLPVPLYIRGYDPTALAEHROTI	1685
Db	2984	GLPKRVVQWVFONARAKEKSKLSMAKHFGINQTSYEGPKTECTLTCIGIKYSARLSVRDHI	3043
QY	1686	-----INDYITSO---QWHEHNTATA---MAOR-ADMLRGLSPRESSIALNLYAAGPR	1729
Db	3044	FSQQHISKVKTOTIGSQLDKKEYFDPAIVRQMAQOELDRI-----KXANEVLCIAQQQ	3098
QY	1730	GIIDLUS--QVPHLPVLVPPTFGTATAMDRLAYLPTAPQPFSSRHSSSPISPG---GPTH	1784
Db	3099	GMFDNTEPLQALNLPTAYPALQGIIPVLLPGL-----NSPSLPGFTPSNTA	3143
QY	1785	LTKPTTTSSSERDRDRDRDREREKSIILSTTVEHAPIWEPGTQSGSSGSSGGG	1844
Db	3144	LTSPPKPLMG-----LFSITT-----VPSFGL-PTSGLPNKESSA	3176
QY	1845	GGSSSRPASHSHAHQHSPTSPTQDALQORPSVLHNTGMKGIITAVEFSPKPTVLRLSTSTS	1904
Db	3177	SLSSTPTPAQTAWA--MGFPQPPQOQQOQQOQOV-----QQPPPP-----	3213

Qy	1905	SPVRPAATFPFATHCPGLGTLGVYPTLMEPVLLPKAEARVAKPERPRADTGHAFIAPKPP	1967
Dd	3214	-----PAAQPPTPOLPL-----QQOQRKDNDSEKVKEKAHKKGEP-LPVFK	3258
Qy	1965	ARGLEPASPSKGSRRPLVPVSGHATTARTPAKNLAPHHA-SDPPAPPASASDPHR	2023
Dd	3259	KEKGEAPTAAITISAPLPW-----EYAVDPAQLQALQAALTSDPTALLTSQFLPYF	3311
Qy	2024	EKTQSKPFSTQEILSRSLGYHGSSYPGVEPVSPVSSPSLTLDKGLPKHLBELDKSHLE	2083
Dd	3312	VPGFSPIYAQIPGALQSGLQPMYGMEGLFFYPSPALSQAL---MGL---	3355
Qy	2084	GELAPKOPGPVKLGCEAANHLPHLRPLPESQPSSSLLOTAPGVKGHORVVTLAQHISEVI	2143
Dd	3356	-----SPGSL-----LQYOQYQOOSLOEAIQ-----QQOORQLQQOQQ-----	3388
Qy	2144	TODVTRHHPOOLGAPLPAPLYSPFGASCPLDLRRLRPDSLYLPPPDHGAPARGSPHSEGG	2203
Dd	3389	-QKVOOQPKASQTTPV-----PGA-----PSPDKD-PAKESPKEEQ	3424
Qy	2204	KRSPE-----PNKTSVLGGEDGI-FPVSPPE-----GNTEFGHSRAV	2241
Dd	3425	KNTPREVSPLLKLPEBFEAESADSLSYDFIVPKVQYKLVCRKCOAGFSDSEAAKSHL	3484
Qy	2242	YPLLYRDGEOTEPSRM-----GSKSPGNTSQPPAFFSKLTESNAMYKS	2286
Dd	3485	KSLCFFQGQSVNVLEQWLHVPTGGGGGGGGGGGGGGGGGGSYHCLACESALCGEAL	3544
Qy	2287	KOEINKKLINTN-----RNEPEVNISQGTETINMPAITGTGLMYRS-----	2329
Dd	3545	SQHLESALHKHRTITRAARNAKEHPSLPHSACFPDEPSTASTSQSAHSNDSPPPPSAAA	3604
Qy	2330	-QAVQEHASTNWGLEAIIRKALMGKYDQWEESPPLSANAFNLINASLASLA-	2379
Dd	3605	PSSASPASHSRKSNQVVSRA3-----RAKPP-----SPFLSSSTVTISSCSCTSGVQ	3652
Qy	2380	-AMPITTAADGRSDHTLTPGQGGGKAKVSG	2407
Dd	3653	PSMPTDDYSSESDTDLQSKDGSPASVPEG	3681
 RESULT 32 TCOF HUMAN			
ID	TCOF HUMAN	STANDARD;	PRT; 1411 AA.
AC	Q13428; Q99408; Q99860;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Treacle protein (Treacher Collins syndrome protein).		
GN	TCOF1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96154183; PubMed=8563749;		
RA	Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K.,		
RA	Bonner C.A., Koprivnikar K., Wasmuth J.J.;		
RT	"Positional cloning of a gene involved in the pathogenesis of		
RT	Treacher Collins syndrome."		
RL	Nat. Genet. 12:130-136(1996).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97228900; PubMed=9074926;		
RA	Dixon J., Edwards S.J., Anderson I., Brass A., Scambler P.J.,		
RA	Dixon M.J.;		
RT	"Identification of the complete coding sequence and genomic		
RT	organization of the Treacher Collins syndrome gene.";		
RL	Genome Res. 7:223-234(1997).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97250498; PubMed=9096354;		

RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,
RA Ashley J.A., Lovett M., Jabs E.W.;
RT "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits
RT mutations in Treacher Collins syndrome throughout its coding
RT region";
RL Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).
RN [4]
RP VARIANTS LEU-439; VAL-810; VAL-1313 AND GLY-1355, AND VARIANT TCS
RX ARG-53.
RY MEDLINE=97195537; PubMed=9042910;
RA Edwards S.J., Gladwin A.J., Dixon M.J.;
RT "The mutational spectrum in Treacher Collins syndrome reveals a
RT predominance of mutations that create a premature-termination
RT codon";
RL Am. J. Hum. Genet. 60:515-524(1997).
CC -!- DISEASE: Defects in TCOF1 are the cause of Treacher Collins
CC syndrome (TCS) [MIM:154500]. TCS is an autosomal dominant disorder
CC of craniofacial development that occurs with an incidence of
CC 1/50,000 live births. The clinical features of TCS are bilaterally
CC symmetrical and include: (1) abnormalities of the external ears,
CC atresia of the external ear canals, and malformation of the middle
CC ear ossicles, which may result in conductive hearing loss; (2)
CC lateral downward sloping of palpebral fissures, frequently with
CC colobomas of the lower eyelids; (3) hypoplasia of the mandible and
CC zygomatic complex; (4) cleft palate.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U40847; AAC50903.1; -
DR EMBL; U76366; AAC51181.1; -
DR EMBL; U84664; AAC51185.1; -
DR EMBL; U84640; AAC51185.1; JOINED.
DR EMBL; U84641; AAC51185.1; JOINED.
DR EMBL; U84642; AAC51185.1; JOINED.
DR EMBL; U84643; AAC51185.1; JOINED.
DR EMBL; U84644; AAC51185.1; JOINED.
DR EMBL; U84645; AAC51185.1; JOINED.
DR EMBL; U84646; AAC51185.1; JOINED.
DR EMBL; U84647; AAC51185.1; JOINED.
DR EMBL; U84648; AAC51185.1; JOINED.
DR EMBL; U84649; AAC51185.1; JOINED.
DR EMBL; U84650; AAC51185.1; JOINED.
DR EMBL; U84651; AAC51185.1; JOINED.
DR EMBL; U84652; AAC51185.1; JOINED.
DR EMBL; U84653; AAC51185.1; JOINED.
DR EMBL; U84654; AAC51185.1; JOINED.
DR EMBL; U84655; AAC51185.1; JOINED.
DR EMBL; U84656; AAC51185.1; JOINED.
DR EMBL; U84657; AAC51185.1; JOINED.
DR EMBL; U84658; AAC51185.1; JOINED.
DR EMBL; U84659; AAC51185.1; JOINED.
DR EMBL; U84660; AAC51185.1; JOINED.
DR EMBL; U84661; AAC51185.1; JOINED.
DR EMBL; U84662; AAC51185.1; JOINED.
DR EMBL; U84663; AAC51185.1; JOINED.
DR EMBL; U79659; AAB40722.1; -
DR EMBL; U79645; AAB40722.1; JOINED.
DR EMBL; U79646; AAB40722.1; JOINED.
DR EMBL; U79647; AAB40722.1; JOINED.
DR EMBL; U79648; AAB40722.1; JOINED.
DR EMBL; U79650; AAB40722.1; JOINED.
DR EMBL; U79651; AAB40722.1; JOINED.
DR EMBL; U79652; AAB40722.1; JOINED.
DR EMBL; U79653; AAB40722.1; JOINED.
DR EMBL; U79654; AAB40722.1; JOINED.

DR EMBL; U79655; AAB40722.1; JOINED.
DR EMBL; U79656; AAB40722.1; JOINED.
DR EMBL; U79657; AAB40722.1; JOINED.
DR EMBL; U79658; AAB40722.1; JOINED.
DR Genew; HGNC:11654; TCOF1.
DR MIM; 606847; -
DR MIM; 154500; -
DR GO; GO:0005730; C:nucleolus; TAS.
DR GO; GO:0005215; F:transporter activity; TAS.
DR GO; GO:0001501; P:skeletal development; TAS.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR003993; treacle.
DR Pfam; PF03546; treacle; 3.
DR PRINTS; PR01503; TREACLE.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS50896; LISH; 1.
KW Disease mutation; Polymorphism.
FT DOMAIN 6 38 LISH.
FT DOMAIN 89 97 POLY-GLU.
FT DOMAIN 204 207 POLY-SER.
FT DOMAIN 616 619 POLY-SER.
FT DOMAIN 919 924 POLY-SER.
FT DOMAIN 1285 1289 POLY-LYS.
FT DOMAIN 1375 1386 POLY-LYS.
FT DOMAIN 1398 1405 POLY-LYS.
FT VARIANT 53 53 W -> R (in TCS).
FT VARIANT 439 439 P -> L (in TCS).
FT VARIANT 810 810 A -> V.
FT VARIANT 1313 1313 A -> V (in dbSNP:15251).
FT VARIANT 1355 1355 D -> G.
FT CONFLICT 1312 1312 /FTId=VAR_005634.
FT SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64;
Query Match 3.0%; Score 391.5; DB 1; Length 1411;
Best Local Similarity 20.4%; Pred. No. 6.1e-06;
Matches 364; Conservative 178; Mismatches 664; Indels 581; Gaps 76;
QY 656 YKRGQNLDEILQHKLMKERNARRKKKAKAPAAAEAEAFPPVVEDEMEA----- 707
DB 50 YTHWQOTSELGRKKAEDALQAKTRVSDPISTESSE-----EEEEAEATAKATPR 104
QY 708 -----SGVSGNE-----EEMVEAEALHASGNEVPGECSGPAT-----VNNSSDTE 749
DB 105 LASTNSSVLGADLPSSMKKAKAEKAGKTGNSMPH-----PATGKTVANLLSGKSPRK 159
QY 750 SI-PSPHTEAAKDTGQNGPKPPATLGADGPPGPPPTPR----- 787
DB 160 SAEPSAMTLVSETEEGSV--AFGA-AAKPGMVSAQADSSSEDSSTSSSDETDVEVKA 216
QY 788 -----RTSRAPLEPTPASEATGATPPPPSPSPSPAPPVVPKKEEKEEETAAAPVBE 840
DB 217 SEKILQVRAASAPAKGTGKAT-----PAPPGKAGAVASQTKAGKEEDESSESSESS 270
QY 841 E-EOKPPAAEEL--AVDTGKAEPEPVKSECTEAEAGPAKGKDAEAAEATAGALKAEKKE 897
DB 271 DSEETPAKALLQAKASGKTSQ--VGAASAPAKESPRKG-----AAPAPP 314
QY 898 GGSGRATTAKSSGAPOSDSSATCSADEVDAAEGDKNRLLSPPSLLTPTGDPANASP 957
DB 315 KTGPAVAKAQAGKREDSQSSESESEEA-----PAQKPSG----- 354
QY 958 QKPLDLKQLKORAAAIPIPIQVTKVHEPPREDAATKPAAPPAPPQNLQPSDAPQPGS 1017
DB 355 -----KAPQVRAASAP-----AKESPRKGAAFA-----PPRTGPAQAQVQVKCQ 394
QY 1018 SPRKSRSPAPPADKEAFAA--EAQ-----KLPGDPCWTSGL--PF-----PVPPREVIK 1064
DB 395 EEDSRSSSESDSDREALAAMNAQVKPLGKSPQVKPASTMGMPGLGKGAGVPPGKGV 454

1065 ASPHA-----PDPSAFSYAPPGLHPLGLHDTARVLPVPRPTISNPPPLISSAKHPSVL 1118
 455 ATPSAQVQKWEEDSESS-----EESDSSDGEVPTAVAP-----AQ 491
 1119 BROIGALISQGNVQLHVPYSEHAKAP--VGPVTMGLPL--PMDPKLAPFSGVKQEQLS 1173
 492 EKSIGNLOAK-----PTSPAPKPPQKAGPVAQVKAEPKPDNSE-----S 533
 1174 PRGAGPESLGVPTAOEASVLRGTALGSVFGSGITKGIPTSTRVPDSALTYRGSITHGT 1233
 534 SEESDSADSEAPAAWTAQAK-----DALKIPQTKACPKKNTTASA 577
 1234 PADVLYKTTIRIIGESPSRLDGRDLSLPGHVIYEGKKGHVLVYEGGMSV-----TQC 1289
 578 KVAPVRVGT-----QPPR-----KAGTATSPAGSSPAVAGTQR 611
 1290 SKEDGRSSSGPPHETAAPKRTYDMWGR--VGRAI-----SSASIEGLMGRALPPERHSP 1342
 612 PAED--SSSEESDEEKTGLAVTVGQAKSVGKGLQVKAASVPVKGSLGGTAPV--- 665
 1343 HHLKEQHIRSGITOGIPRSVYEAQEDVLRREKLLKREGTPPPPPSRDLTEAYKTOAL 1402
 666 -----LPGKTGPTVQVKAQED--SESSEESDSDEAAASPAQVKTSVKKTQA- 713
 1403 GPLKLPKPAHEGLVATVKEAGRSIHEI PRELRHTPELPAPRLPKESITQCTPLKYDTG 1462
 714 ---KANPA-----AARAPAKGTIS--APGKVTA 738
 1463 ASTTGSKKHVRSILIGSPRTPPVH-----PLDVMADARALERACYEESLKGR 1511
 739 AAQAKR-----SPSKVPPVFNQNSVTLARGPASVPSVKAATAA-----QAQ 784
 1512 POTASSSGSIARGAPVIVPE-LGKPROSLTYEDHG--APFAGHLPGSPVTRRETPR 1568
 785 TGPEEDSGSSEESDEEAEETLAQKPSGKTHQIRAAALP-----AKESPRKGAAPTTP 839
 1569 LQEGSLSSKASQDRKLTSTPREIAKSPHSTVPEHHPHPISYEHLLRGVSGVDLYRSHI 1628
 840 GKTGSPAAQKQDDDS-GSSSEESDSDEAPAAVTSQVIRP----- 880
 1629 PLAF-DPTSIPIRG--IPLDAAAAYLPHRLAPNPTYPHLYPPVLYRGPDTAALENROT 1684
 881 PLIFVDPNPSAGPAATPAQAQA-----STPRKARASESTARSSES 924
 1685 INDIYTSQ-----MHNTAT-----AMAQADMLRGLSPRESSLAIN--YAAGPRGI 1731
 925 EDEDVIPATQCLTPGIRTNVVTMTAHPRIAPKASWAGASSKESRISDGKKQEGPATQ 984
 1732 IDLSQVPHLPVLVPPPTPGTPATAMDRLAYLPTAQPFFSR--HSSSPLSPGPGTHLTKPT 1789
 985 VSKGNPASLPL-----TQAALKVLAQKASEAQPVPVARTQPSGSDVAVGTLPATSPQ 1036
 1790 TTSSSRERDRDRDRERERKSIITSTTVEHAPIWPGTEQSGSGSGSGSGSGSGSS 1849
 1037 STSVQAKGTINKLR-----KPKLPEVQATKAP--ESSDDSDSDSDSSGSEEDGE 1084
 1850 RPASHSHAHQHSPISPRTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSPVRP 1909
 1085 GPOGAKSAHLTGTPBRTETLVEE-----TAASSEDDVV----- 1119
 1910 AATFPFATHPLGGTLDGVVYPTLMERFVLLPKAPRVARPERPRADTGHFLAKPARSGL 1969
 1120 -----APSQSLLSG---YMTFGLTP--ANSQASKATPKLD-----SSPSVSSTL 1158
 1970 EPASSPSKGEPRP-----LVPPVSG-HATYARTAKNALPHASPDPPAPASADPH 2022
 1159 AAKDDPDGKQEAQPOQAAGMLSPKTKGKEAAGSTTPKSRKKPKKAGNFOASTLAQSMI 1218
 2023 REKTQSPFSGIOELRLSLGHSYSGVEGVPSPVSPSLTHDKLPLKHELDKSHL 2082
 1219 TQCLLGQFWPNEAQA-----SVMKVLTTEL-----L 1246

2083 EGELRPKQPGPVKLGGEAAHPLHLPLPESPFSSSPLLOTAPGVKHORVVTLAQHISEV 2142
 1247 EQE-RKKVDDTTKSSRKGWESRKRKLGQDPA--RTPRSKKKKL----- 1290
 2143 ITQDTRIHPHPOOLSAPLPAPLYSPGASCPVLDLRRPFDLYLPPDHGAPAGSPHSEG 2202
 1291 -----GEG 1295
 2203 GKRSPEPNKTSVLGGEDGIEPVSPPEGMTEPHGSRSAVYPLLYRDGQTEFSR---M 2257
 1296 GEASVPEKTSITSTGKAKDKAS--GDVKEKKGKS---LGSQAKDEPEEELQKGM 1348
 2258 GSKSPGNTSOPPAFPKSLTESAMVSKKQKQIN--KKLNTNRNEP 2302
 1349 GTVEGDDQSNPKSKKSKKDRKKDKKKEKKKKAKKASTKXDSBP 1395
 RESULT 33
 CALL CANFA
 ID CALL CANFA STANDARD; PRT; 1460 AA.
 AC OQXSJ7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT OI ALA-208.
 RC TISSUE=Skin;
 RX MEDLINE=21023337; PubMed=11147834;
 RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
 RT "Sequence of normal canine COL1A1 cDNA and identification of a
 RT heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case
 RT of canine osteogenesis imperfecta.";
 RL Arch. Biochem. Biophys. 384:37-46(2000).
 CC -!- FUNCTION: Type I collagen is a member of group I collagen
 CC (fibrillar forming collagen).
 CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -!- PM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
 CC (OI).
 CC -!- SIMILARITY: Contains 1 VWFC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC EMBL; AF153082; AAD34619.1;
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR InterPro; IPR001007; VWFC.
 DR Pfam; PF01410; COLF1; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; C1g_helix; 2.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLF1; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWFC_1; 1.
 DR PROSITE; PS0184; VWFC_2; 1.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal; Disease mutation.
 FT SIGNAL 1 22 BY SIMILARITY.

FT PROPEP 23 157 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 158 1214 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1215 1460 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 34 92 WFPC.
FT DOMAIN 158 174 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 175 1188 TRIPLE-HELICAL REGION.
FT DOMAIN 1189 1214 NONHELICAL REGION (C-TERMINAL).
FT SITE 741 743 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1089 1091 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 1361 1361 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARIANT 208 208 G -> A (in OI: severe).
SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 3.0%; Score 391.5; DB 1; Length 1460;
Best Local Similarity 21.4%; Pred. No. 6.3e-06;
Matches 379; Conservative 102; Mismatches 645; Indels 643; Gaps 82;

Qy 722 BALHASNEVPRGCSGATVNNSSDTSIESPHTAAKDTGONGPKPPA-TLGDGPP- 779
Db 73 ETKNCPGAQVPPGCCPCVCPDGEASPTDQETGVGPKGDTGPRGPGAGPPGCRDGI 132
Qy 780 ----PGPTPRRTSRAPIEPTPASEATGATPP-----GPPGPGLGNGFAPQMSYGVDKSTGSI 811
Db 133 QPGLPGPPCP-----GPPGPGLGNGFAPQMSYGVDKSTGSI 175
Qy 812 PPSPSAP-----PPVVPKEEKEETAAAPPVEEG-----EOKPPAAEELAVDTGKA 858
Db 176 PMGSPGRGLPGPGAPGPGQFQ---GPPGPGEGAGSGPMGPPGPPGPKNGDGEA 231
Qy 859 EEPVKSECTEAEAGPAKDAEAAEATAEAGLAKAEKKEGSGRATTAKSSGAQDSDSS 918
Db 232 GKPR-----PGERGPPGQARGLPGTA--GLPMKGHRGFGSLDGAKGADGP- 278
Qy 919 ATCSADEVDEAGDKNLLSP----RPSLLTPTGDPANASPOKPLDLKQKQRAAIP 974
Db 279 ----AGKGEPPGNGAPGQMGPRGLPGERGRGAP-----GPAGARG 319
Qy 975 PIQVTKVHEPPREDAAPTKA-PPAPPPQNLPESDAPQPGS--SPRGKSRSPAPP--- 1029
Db 320 NDGATGAAGPP---GPTGAPGPGFPGAVGAKGAGQAGSGEGPGVGEFGPGGPA 375
Qy 1030 ----ADKAFABAOKLPDPPCWTSGLP-FVPPPREVIKASHPADPSAFSIA 1078
Db 376 GAAGPAGNPGADGQFAGKANGAPG-----IAGAPFGFARGSPGSPGPPGPKNSG 430
Qy 1079 PPGHPLPLGLHDTARVLPRTTISNPPPLISSAKHPSVLROJCAISQMSVQLHPYS 1138
Db 431 EPAGPKNKG--DTGAKGEPGTGIGPP-----GPAGEGKRG----- 466
Qy 1139 EHAKAPVGPVTWGLPLPMDPKLAPFGVKOEQLSPRQAGPPSPSLGVPVTAQEA 1198
Db 467 --ARGEFGPT--GLP-----GPPGERGPGSRGFPFGADGVAGPKGP 503
Qy 1199 A--LGSVPGSGITKIP-STRVPSDSALTIRGSIHTGTPADVLKGTITRIIGBDS 1255
Db 504 AGERGS--PGAPGKPSGCEAGRPGEAGLPGAKGLT-GSPGSPGPDGK----TGPPGAGQ 557
Qy 1256 D-----RGEDSL-----PKGHVYEGKGVLSYEGM-----SVTQCSKEDGR 1295
Db 558 DGRGPPGPGARGQAGVMGPPGKGAAGEPGKAG-----ERGVGPPGAVGAPGKGEA 612
Qy 1296 SSSGPPHETAPKRTYDMWGRVGRASIASIEGLMGRAPP-ERHSPHHLKEQHIRGS 1354
Db 613 GAQGGP--GPAGPAGE---RGEQGA-GSPGFGQLPGPAGPPGAGKGE----- 656
Qy 1355 ITQGIPIYSYEAQEDYLRRREAKLKRKGTTPPPPPSRDLTEAYKQALGPLK-L-PAHE 1412
Db 657 --QGVPGDLGAPGSGARGRGPGRGVQVPPGPA-----GPRGANGAPGND 702
Qy 1413 GLVATVKEAGHSIHEIFREELRHFTPELPLAPRLKEGSIQTGTP-----LKVDTCAS 1468
Db 703 GAKG---DAG-----APGAP-----GS--QCAPGLQMPGERGAAGLPG 736

RESULT 34

ANK2 HUMAN

ID ANK2 HUMAN STANDARD; PRT; 3924 AA.

AC Q01484; Q01485;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).

Qy 1469 KKHD-----VRLJGSPGR-----TFP--PVHPLDVMADARALARACYEESLSKRPCTA 1515
Db 737 PKGDRGDAGPGADGSPGKGVRLGTGIPGPPAGAPGD-----KGEAGPS 783
Qy 1516 SSSGSGTARGAPVIVPELGKPRQSPPLYTHEDGAPFAGHLPRGSPVTMBEPTPLRQEGSL 1575
Db 784 GPAGTARGAPGDRGEPGP--GPAGFA--GPPGAD---GQPGAKGEPGDAGAKGD-- 833
Qy 1576 SSKASQDKLTSTPREIAKSPHSTVPEHHPI-----SPYEHLLRGVGVLDLYRSHIPLA 1631
Db 834 -----AGPCPAGCTCPFPPIGNVGAPGPKGARGSAGP----- 866
Qy 1632 FDTSPISRGIFLDAAYYLPHRLAPNTYPHLYPYLIRGYDPTAALENRQTIINDYIT 1691
Db 867 -----PCATGPGCAAGRVG--PGPS--GNAGPP-----GPPGAKGEGKGARGE-- 908
Qy 1692 SQQHHNTATAMAQADML-----RGLSPRESSLALNYAAGPRGIIDLQVPHL 1740
Db 909 -----TGPRGRPGEVGPPGPPGAGEKSGPGADGAPGTPGPGIAGQGVVGL 959
Qy 1741 PVLVPT--PGTPATAMDRLAYLTPAPOPFSRSHSSPLSPGPGTHLTPTTTSSSERER 1798
Db 960 PQQGERGFGGLPGSGE-----PKQGPSSTSGERGPPGMPGCLAGPPGEGS-- 1009
Qy 1799 DRDRDRDREREKSIILTSITTTVEHAPIWRPECTEQQSSGSSG-----GGGSSSRPAS 1853
Db 1010 -----REGS-----PCAEGSPGRDGSPPGKGRGETGAPGPGA 1043
Qy 1854 HSHAQHSPISTPTQDALQORPSVLNWTMGKGIITAVEPSKPTVLSTSTSPVTPAATF 1913
Db 1044 PGAPCAPGVPG-----AGKNG-----DRGETGAPGAG-- 1072
Qy 1914 PPATHCLGGTLDGVYPTLMEPVLLPKEAPRVARPERPRADTGH-----A 1958
Db 1073 -----FIG-----PVGARGPAGQGRGKGETGEQDGRGKRGHGF 1110
Qy 1959 FLAKPPARSGLEPASSPSKSGSEPRPLVPPVPSGHATARTPAKNLAPHHASPDPPAPASA 2018
Db 1111 GLQGGPPGPGSPGEGPSGASGP-----AGPRGPPGSA 1143
Qy 2019 SDPHREKTQSPFSIOELLELSLGYHSSYSPEGVFVSPVSSPSLTHDKGLPKHLELD 2078
Db 1144 GSPGKGLNGLPPI-----GPPGPRGRTGAGVPVGPFGPG-- 1180
Qy 2079 KSHLEGLRKPQPGVKLGGEAAHLPHLRPLPEQSPSSPLLQTAGVKHQ----- 2130
Db 1181 -----PPGPPGPPSGGPFDFLP-----PQKADGGYYRAD 1215
Qy 2131 -----RVVTLAQHSIVITQDYTRHHPOQLSAPLPAPLYSFFCASCPLDL 2176
Db 1216 DANVVRDRDLEVDITLKSLSQIENIRSPGSRKNPAR-----TC--RDL 1258
Qy 2177 RRPDSLYLPPDHGAPARGSPHSEGGKRSPEPKNTSVLGG-----CEDGLEPVP 2227
Db 1259 KMCHSDW-----KSGEYWTDPNQCNCNLDKIVFCNMNETGETCYVTPQ 1301
Qy 2228 PEGM-----TEPGHSRSVAVPLLYRDEQTEPSPRMSKSPGNTSQPPAFPSKLTESNSA 2281
Db 1302 QVAKNWIYISKNPKKHHVYGESMTDGFQFEG-----GQSGDPAADVAIQLT-- 1349
Qy 2282 MVKSKQEIKNKLNTHNRNEPEYNISQFG 2310
Db 1350 FLRLMSTEASQNIYTHCKNSVAYMDQQTG 1378

GN ANK2.
OS Homo sapiens (Human) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBT_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE=Brain stem;
RX MEDLINE=91302466; PubMed=1830053;
RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT "Isolation and characterization of cDNAs encoding human brain
RT ankryns reveal a family of alternatively spliced genes.";
RL J. Cell Biol. 114:241-253(1991).
RN [2]
RP REVISIONS.
RA Carpenter S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain stem;
RX MEDLINE=94075409; PubMed=8253844;
RA Chan W., Kordeli E., Bennett V.;
RT "440-kD ankrynB: structure of the major developmentally regulated
RT domain and selective localization in unmyelinated axons.";
RL J. Cell Biol. 123:1463-1473(1993).
RN [4]
RP SEQUENCE OF 463-495 FROM N.A.
RX MEDLINE=92009921; PubMed=1833308;
RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA Lux S.E., Ward D.C., Forget B.G.;
RT "Isolation and chromosomal localization of a novel nonerythroid
RT ankryn gene.";
RL Genomics 10:858-866(1991).
CC -!- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC elements. Also bind to cytoskeletal proteins.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q01484-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q01484-2; Sequence=VSP_000267, VSP_000268;
CC Name=3;
CC IsoId=Q01484-3; Sequence=VSP_000268;
CC -!- TISSUE SPECIFICITY: Plasma membrane of neurons as well as glial
CC cells throughout the brain.
CC -!- PTM: Phosphorylated at multiple sites by different protein kinases
CC and each phosphorylation event regulates the protein's structure
CC and function (Potential).
CC -!- SIMILARITY: Contains 23 ANK repeats.
CC -!- SIMILARITY: Contains 1 death domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi-sib.ch/announce/>
CC or send an email to license@ebi-sib.ch).
CC -----
DR EMBL; X56957; CAA40278.1; -;
DR EMBL; X56958; CAA40279.2; -;
DR EMBL; Z26634; CAB42644.1; -;
DR EMBL; M37123; AAA62828.1; -;
DR PIR; S37431; S37431.
DR HSSP; P42771; 1DC2.
DR Genew; HGNC:493; ANK2.
DR MIM; 106410; -;
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000488; Death.
DR InterPro; IPR000906; ZUS.
DR Pfam; PF00023; ank; 24.
DR Pfam; PF00531; death; 1.
DR Pfam; PF00791; ZUS; 1.

DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 22.
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00218; ZUS; 1.
DR PROSITE; PS50088; ANK REPEAT; 20.
DR PROSITE; PS50297; ANK REP REGION; 1.
DR PROSITE; PS50017; DEATH DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW Phosphorylation.
FT REPEAT 63 92 ANK 1.
FT REPEAT 96 125 ANK 2.
FT REPEAT 129 158 ANK 3.
FT REPEAT 162 191 ANK 4.
FT REPEAT 193 220 ANK 5.
FT REPEAT 232 261 ANK 6.
FT REPEAT 265 294 ANK 7.
FT REPEAT 298 327 ANK 8.
FT REPEAT 331 360 ANK 9.
FT REPEAT 364 393 ANK 10.
FT REPEAT 397 426 ANK 11.
FT REPEAT 430 459 ANK 12.
FT REPEAT 463 492 ANK 13.
FT REPEAT 496 525 ANK 14.
FT REPEAT 529 558 ANK 15.
FT REPEAT 562 591 ANK 16.
FT REPEAT 595 624 ANK 17.
FT REPEAT 628 657 ANK 18.
FT REPEAT 661 690 ANK 19.
FT REPEAT 694 723 ANK 20.
FT REPEAT 727 756 ANK 21.
FT REPEAT 760 789 ANK 22.
FT REPEAT 793 822 ANK 23.
FT DOMAIN 1773 1950 REPEAT-RICH REGION.
FT REPEAT 1773 1784 REPEAT A.
FT REPEAT 1785 1796 REPEAT A.
FT REPEAT 1797 1808 REPEAT A.
FT REPEAT 1809 1820 REPEAT A.
FT REPEAT 1821 1832 REPEAT A.
FT REPEAT 1833 1844 REPEAT A.
FT REPEAT 1845 1856 REPEAT A. (APPROXIMATE).
FT REPEAT 1857 1867 REPEAT A.
FT REPEAT 1868 1879 REPEAT A.
FT REPEAT 1880 1891 REPEAT A.
FT REPEAT 1892 1902 REPEAT A. (APPROXIMATE).
FT REPEAT 1903 1914 REPEAT A.
FT REPEAT 1915 1926 REPEAT A.
FT REPEAT 1927 1938 REPEAT A.
FT REPEAT 1939 1950 REPEAT A.
FT DOMAIN 3536 3620 DEATH.
FT VARSPPLIC 1039 1039 Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).
FT I -> S (IN REF. 1).
FT QY -> HA (IN REF. 1).
FT I -> Y (IN REF. 1).
FT Q -> OFLGKHLPTAPPPLNKGSLVSRILQLGPPGTK
FT (in isoform 2).
FT /FTId=VSP_000267.
FT Missing (in isoform 2 and isoform 3).
FT /FTId=VSP_000268.
FT Q -> PE (IN REF. 4).

Db 943 NSYRLSWGTELDNVALSSPIHSGFLVIFWMDARGGAMRGCRHGL-RIIIPPRCKTAP 1001
Qy 240 A-----AHRI-----LEG-----LGPQ-VELPLYNQPSDTRQYH 267
Db 1002 TRVTCRLVKRHLATMPWVEGEGLASRLIEVGPSCAQLFVIVEIHPH----- 1051
Qy 268 ENIKINQAMRKKLILYFKRNRHAKQWKQFCQRYDQLEALEKKVKVRIENNRRAKES 327
Db 1052 -----AALRGKERELVLRSENGDSWKEHFCDYDELSLNLGMDVLDSPEDLEKK- 1104
Qy 328 KVREYEQPFEIRKQRELQRMQSRVQORGSLMSAARSEHEV-----SEIIDGLS 380
Db 1105 RICRIITDFPOYFAVSRIRKQDNLGPEGVLSSTVVPQVAFPEGALTKRIRVGLQ 1164
Qy 381 EQ-----ENLEKQMRQLAVIPMLVADQORIKF-----INNGMLMAD-- 418
Db 1165 AQPHESELVKILGNKATFPIV-TLEPRRRKFHKPITMTIPVPKASDDVLMNGFGDAP 1223
Qy 419 -----PMKYKO-----ROVMNMWSEQEK 437
Db 1224 TLRLCSITGGTTPAQWEDITGTTPLTFVNECVSFTTNVSARFMLIDCKQI-----QES 1277
Qy 438 ETR-----EKFMQHPNFGLIA-- 455
Db 1278 VTFASQVYREIICVPYMAKFVFAKSHDPIEARLRCFCWTDKDKTLEQQENFAEVARS 1337
Qy 456 --SFLERKTIV-AEC-----VLYYLTKNE-----NYKSLVRSY 487
Db 1338 RDVEVLEKPIYVDCFGNLVPLTKSGQHIIFFFAFKENRLPLFVKVRDRTTQEPGCRLSF 1397
Qy 488 RRRGKSQQQQQQQQQQQQQQMPRSSQBEKDEKEKEAEEKEK-PEVENDKEDLL 546
Db 1398 MKEPKS---TRGLVHQALCNLNLITLPIVTKESQEQEEDIDTSEKNDTESTETSVL 1454
Qy 547 KE-----KTDDTSQDNDKEAVASKGRKTANSQGRKGRITRSMANEANSEERAIT 597
Db 1455 KSHLVNEVPVLASPDLLSEVEMKQDLIKMTAILTTDVSADKAGSIKVKELVKAABEEEPGE 1514
Qy 598 PQQSAELASMEI---NE---SSRWTEBEMET-----AKGILLEHGRNNSAIAIRMVGSKTVS 647
Db 1515 PFEIVERKEDEKNEILRSGTCTRDSSVQSRSEGLVE--BEWVIVS----- 1563
Qy 648 QCKNFYNYKKRONLDELIOQHLKMEKERNARRKKKAPAAASFEAAPPVWE---DEE 704
Db 1564 -----DEBIEE-----ARQKAPLEITEY---FCVEVRIDKE 1591
Qy 705 MEASGVSGNEEMVEBAAHALHASGNEVPRGCSGPATVNNSSD--TESIPSPHTEAAKDT 762
Db 1592 IKG-----KVEKDSGTGLVNYLTDLLNTCVLPKBLQQLQTVQ 1626
Qy 763 GONGPKPPA-TLGADGPPPGPPTPRRTSRAPIBPTPASEATGAPTPPPAPSPSAPPV 821
Db 1627 DKAKKCEALAVGRSSEKGGKDIIPDETQ-----STQKHKPSLIGIKKPV 1671
Qy 822 ---VPKEEBEETAAAPPVEBGEQKPPAAEBELAVDTGKAEBPVKS----- 864
Db 1672 RRKLKQKQKEGLQASAEKELKGSSEESLGEDPGLAPELPTVKATSPLEETPIG 1731
Qy 865 -----ECTEAEEGPA-----KGDAEAAEAT-----ABGALKAEKKEGGSGR 902
Db 1732 SIKDKVKALQKRVDEQGRSKLPIRVKGBDVPKKTTHRPHAPASPSLKSERHAPGSPS 1791
Qy 903 ATTAKSGAPQDSSATCSADEVDEAGBGDKNLLSPRSLTPTGDPANAS---PQK 959
Db 1792 PKTERHSTL-----SSAKTERHPVSPSKTEKHSFVS--PSAKTERHSPASSSKTEKHS 1846
Qy 960 PLDLKQLKQRAAATPIPIQVTKV-HEPPREDAAPTKPAAPPAPPQNLQOPESDAPQOQCSS 1018
Db 1847 PVSSTKTERHS---PVSSTKTERHPVPSPGKTDKRPVPSGRT---EKHPPVPSGRT 1900
Qy 1019 PRGKRSPPADKFAFAEAQK-----LPGDPPCWTGSLPPPPVPP-----RE 1061

Db 1901 EKRLVPSPSGRTDKHQPVSTAGTKTEKHLVPSPSGKTEKOP-PVSPSTKTERIEETMSVRE 1959
Qy 1062 VIKASPHAPDPSAFSAAPPGLPLGLHDHTARVLPVRPRTTINPPPLISSAKHPSVLERQ 1121
Db 1960 LMKAFOSQODPSK-----HKTGLFEHKSQKQPO-----EKGVVRVEKEGKPILTQRE 2008
Qy 1122 IGASIQGMSVOLHVPISEHAKAPGVPTWGLPLPMDPKKLAPFSGVKQEQSLSPRQAGPP 1181
Db 2009 -----AQKTENQTIKRGOR--- 2022
Qy 1182 ESLGVPTAQEASVLRTALGALVPGGSIYTKGIPSTRVPSPDSALTIRGSIHTGTPADVLKYG 1241
Db 2023 ----LPVTGTAEKRGVRVSSI-----GVK----- 2043
Qy 1242 TITRIIGEDSPSLDRGREDSLPKGHVIEYEGKHGHLVSEYEGMSVTCQSKEDGRSSGPP 1301
Db 2044 -----KEDA-----AGKEKVLSHKIPPEPVQSVPEEESHRESEVP 2078
Qy 1302 HETAPKRTYDMMGRVGRVGRASSASIEGLMGRAIPPERHSPHHLKE-----QHHRIGST 1356
Db 2079 KERMADEQDMDLQIISPDKTSTDFSEVIKQBELENDKYQOFRLESETEKAQLHLDQVLT 2138
Qy 1357 Q-----GIPRSYVE-----AQEDYLREAKLLKREGTTPPP-----PPSRDLTEAY 1397
Db 2139 SPFNTPPLDYMKDEFPLALSLQSGALDGSSESLKNEGVAGSPCGSLMEGTPOISSESY 2198
Qy 1398 KTOALGPLKAPAEGLVATVKEAGRSITHEIPRELRHTPELPLAPRLPKEGSIQTGTPL 1457
Db 2199 K-----HEGLAETPETSLSFSPKKSSEOTGE-----TKESTKTETTT- 2238
Qy 1458 KYDTGASTGSKKHVRSLSIGPGRTPFPVPHPLDVMDARALERACYEESLKSRRPGTASS 1517
Db 2239 -----EIR-----SEKEHTTKDI 2252
Qy 1518 SGGSIARGAPVIVPELKGPROSLTYEDHGAPFAGHLPRGSPVTVNREBTPRLOE---GSL 1574
Db 2253 TGSSEERGATVEDS-----ETSTESFOKEATLGSP---KDTSPKRQDDCTGSC 2298
Qy 1575 SSSKA-----SODRKLTSTPREIAKSPHSTVPEHHPHPIPSYEHLLRGVSGVDLYRSH 1627
Db 2299 SVALAKETPTGLTEAACDEGQRTFGSAHKTQDSE----- 2335
Qy 1628 IPLAFDPTSPRGIPIDAAAAYILPHRLAPNPTYPHLYPPYLIRGYPD--TAALENRQTI 1685
Db 2336 ---AQESTATS---DETALPPEASVKTDTGTESKPGQVIRS-PQGLELALPSRDS- 2385
Qy 1686 INDYITSQOMHNTATAMQADMLRGLSPRESSIALNY-----AAGPGIIDLQVPHLP 1741
Db 2386 -----EVLASA-----DDSLAVSHKDSLEASP--VLEDNSSHKTP 2419
Qy 1742 VLVPTPTCTPATMDRLAYLPTAPQPFSSRHSSSPLSP---GG--PTHLTPTTTSSSR 1796
Db 2420 DLSLPSPLKESPCRDLSL-----ESSPVEPKMAGIPPHFPLPAAVAKTEL 2465
Qy 1797 ERD-----RDR-ERDRDREREKSIILTTTTVBAPHIWRPGTEQSGSGSSGGSGSSSRP 1851
Db 2466 LTVASVRSRLRRDPDGSAAEDDSLEQTSMB-----SSGKSPLSPDPT 2508
Qy 1852 ASHSHAHQHSPISPRTQDALQOORFVSLHNTG-----MKGIITAVE-- 1891
Db 2509 SSEEVSVE---VTPKTTDVTSPKPAVITHECAEBDDSENKEKRTPEEEMFMVTKIMF 2565
Qy 1892 -----PSKPTVLRSTSTSSPVRPAAT-FPPATHCPLGGTLDGVVPTLMEPVL 1937
Db 2566 DELEQBAKQKDYKKEPKQEESSSSDDDCSDVDVDEPKTGTGDESGV-----PVL 2619
Qy 1938 LPKEAPRV-----ARPE-----RPRADTGHAFIAPK-----PARSGLEPASSPSKSGE 1980
Db 2620 VTSESRKVSSSESEPELAQLKKGADSG--LLPEPVRVQPPSPPLPSMDSNSP---E 2673
Qy 1981 PRPLVPPVSGHATTARTPAKNLAPHHASPPDPAPASADPHREKTKQSKPFSIOEL----- 2036
Db 2674 EVQFPVVSKQYTPKMNEDTQEEEPKSEESKDSHLEADRHAVSTEADRSYDKLNRT 2733

QY 2037 -ELSLGHHSSVSGVEPVSS---PVSS-----PSL-----THDKGL 2070
Db 2734 DQPKICDHGC-----EAMSPSSARPVSSGLQSGPTGDVDEQPVYKESLALQTHEKDT 2789
QY 2071 PKHLEEDKHLGELRPGKQPGVKLGGEAAHPLH-----RPLPESQPSSSPLLQAPG 2125
Db 2790 EG--EELDVRAS--PQADCSSESFSSSSPLHCLVSGKELDEDISATSIQKTE-- 2842
QY 2126 VKGHRVVTIAQHISEVITQYTHHPQQLSAPLPAPLYFPQASCPVLRLRPPSDLYL 2185
Db 2843 -----VTKTDETFENLPKDCP-----SQDSSITQTDRFSM---DVPVSDLAENDBIYD 2888
QY 2186 P---PPHGPARGSPHSEGGKSPENKTSVLGGGE-DGIEPVSPPEGMTEPHGSRSAV 2241
Db 2889 PQITSPYENVPQSOFSSSESKTQTDANHTTSPHSSSEVYVTTTSPVEDVVVASSSGTV 2948
QY 2242 YPLLYRDG-----EOTEPS--RMGSKSPGNTSQPPA----- 2270
Db 2949 ---LSKESNEGQDIKMESQLESTLWEMQSDVSQSSFEPTMSATTTVVGEQISKVIITKT 3005
QY 2271 -----PFSSKLTESNSAM---VSKKKQEI-----NKKLTHNRNEPEYNISQPGT----- 2311
Db 3006 DVDSDSWSEIRDEDEAFEARVKEBEQKIFGLMVDROSGQITPTDTPARTPEEGTPTSEQ 3065
QY 2312 -----EIFNPAITGTLMTYRSQAVQEHASTNMGLERAIIRKALMGKYDQWESPPPL 2363
Db 3066 NPFLFQEGKLFEMTR-SGAIDMTKRSYADESFHFQIQES-----REETLSEDKVE 3116
QY 2364 SANAFNPLNASALPA-----AMPITAAD-----GRSDHTLTSPGGGKA 2403
Db 3117 GATGADPLPLETSAESLALSBSKETVDDEALLPDSVSEVEEIPASDAQLSQMG--I 3173
QY 2404 KVGSRPSRKAKS-----PAPGLAGDRPP-----SVSSVISEGDCNRRPTLTN 2447
Db 3174 SASSTETPKAVSVGTGDLPTVQTGDIPLPLSGVKQISCPDSSEFAVQQLDF---SLTR 3230
QY 2448 RVWEDR---PSSA-----GSTPPPYPLMLRQAGVGNVMSPPPPGLPAGS 2488
Db 3231 SVYSDRGDDSPDSPEEGKSVIEPTAPMENVPFTESKSKIPVMTPTSTPAPSAEVES 3290
QY 2489 GPLAGPHAWDEEPK 2503
Db 3291 SVSEDFLSSVDEENK 3305

RESULT 35
CA17 HUMAN
ID CA17 HUMAN STANDARD; PRT: 2944 AA.
AC Q02388; Q14054; Q16507;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(VII) chain precursor (Long-chain collagen) (LC collagen).
GN COL7A1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94327588; PubMed=8051117;
RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;
RT "Cloning of human type VII collagen. Complete primary sequence of the alpha 1(VII) chain and identification of intragenic polymorphisms."; J. Biol. Chem. 269:20256-20262(1994).
RN [2]
RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93338437; PubMed=1307247;
RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G., Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E., Uitto J.

RT "The large non-collagenous domain (NC-1) of type VII collagen is amino-terminal and chimeric. Homology to cartilage matrix protein, the type III domains of fibronectin and the A domains of von Willebrand factor."; Hum. Mol. Genet. 1:475-481(1992).
RL [3]
RN SEQUENCE OF 815-1439 FROM N.A.
RP MEDLINE=91334380; PubMed=1871109;
RA Parente M.G., Chung L.C., Ryyanaenen J., Woodley D.T., Wynn K.W., Bauer E.A., Mattei M.-G., Chu M.-L., Uitto J.;
RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the gene."; Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).
RL [4]
RN SEQUENCE OF 369-1255 FROM N.A.
RP MEDLINE=93107742; PubMed=1469284;
RA Gammon W.R., Abernethy M.L., Padilla K.M., Prisyavanh P.S., Cook M.E., Wright J., Briggman R.A., Hunt S.W. III;
RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain extracellular matrix."; J. Invest. Dermatol. 99:691-696(1992).
RL [5]
RN SEQUENCE OF 340-675 FROM N.A.
RP TISSUE=Keratinocytes;
RC MEDLINE=92231902; PubMed=1567409;
RA Tanaka T., Takahashi K., Furukawa F., Imamura S.;
RT "Molecular cloning and characterization of type VII collagen cDNA."; Biochem. Biophys. Res. Commun. 183:958-963(1992).
RL [6]
RN SEQUENCE OF 2395-2944 FROM N.A.
RP MEDLINE=93271985; PubMed=8499916;
RA Greenspan D.S.;
RT "The carboxyl-terminal half of type VII collagen, including the non-collagenous NC-2 domain and intron/exon organization of the corresponding region of the COL7A1 gene."; Hum. Mol. Genet. 2:273-278(1993).
RL [7]
RN SEQUENCE OF 1-87 FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=94375010; PubMed=8088784;
RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W., Uitto J., Greenspan D.S.;
RT "Structural organization of the human type VII collagen gene (COL7A1), composed of more exons than any previously characterized gene."; Genomics 21:169-179(1994).
RL [8]
RN VARIANT BDDSC ARG-2034.
RP MEDLINE=89227237; PubMed=2653224;
RA Fine J.D., Johnson L., Wright T.;
RT "Epidermolysis bullosa simplex superficialis. A new variant of epidermolysis bullosa characterized by subcorneal skin cleavage mimicking peeling skin syndrome."; Arch. Dermatol. 125:633-638(1989).
RN [9]
RP REVIEW ON DEB VARIANTS.
RX MEDLINE=98041696; PubMed=9375848;
RA Jaervikallio A., Pulkkinen L., Uitto J.;
RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in the type VII collagen gene (COL7A1)."; Hum. Mutat. 10:338-347(1997).
RL [10]
RN VARIANT DEB LYS-2798.
RX MEDLINE=93291877; PubMed=8513326;
RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y., Lin A.N., Dietz H.C., Hovnanian A., Uitto J.;
RT "A missense mutation in type VII collagen in two affected siblings with recessive dystrophic epidermolysis bullosa."; Nat. Genet. 4:62-66(1993).
RL [11]
RN VARIANT DEB SER-2040.
RX MEDLINE=94224777; PubMed=8170945;
RA Christiano A.M., Ryyanaenen M., Uitto J.

RT "Dominant dystrophic epidermolysis bullosa: identification of a
RT Gly-->Ser substitution in the triple-helical domain of type VII
RT collagen.";
RN Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553 (1994).
RN [12]
RP VARIANT DEB CYS-2623.
RX MEDLINE=96081220; PubMed=8541842;
RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Uitto J.;
RT "Pretibial epidermolysis bullosa: genetic linkage to COL7A1 and
RT identification of a glycine-to-cysteine substitution in the triple-
RT helical domain of type VII collagen.";
RN Hum. Mol. Genet. 4:1579-1583 (1995).
RN [13]
RP VARIANT DEB ARG-2043.
RX MEDLINE=95164985; PubMed=7861014;
RA Christiano A.M., Morriconi A., Paradisi M., Angelo C., Mazzanti C.,
RA Cavallieri R., Uitto J.;
RT "A glycine-to-arginine substitution in the triple-helical domain of
RT type VII collagen in a family with dominant dystrophic epidermolysis
RT bullosa.";
RN J. Invest. Dermatol. 104:438-440 (1995).
RN [14]
RP VARIANTS DEB
RX MEDLINE=96220218; PubMed=8644729;
RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;
RT "Glycine substitutions in the triple-helical region of type VII
RT collagen result in a spectrum of dystrophic epidermolysis bullosa
RT phenotypes and patterns of inheritance.";
RN Am. J. Hum. Genet. 58:671-681 (1996).
RN [15]
RP VARIANT DEB ARG-2575.
RX MEDLINE=96154068; PubMed=8592061;
RA Shimizu H., McGrath J.A., Christiano A.M., Nishikawa T., Uitto J.;
RT "Molecular basis of recessive dystrophic epidermolysis bullosa:
RT genotype/phenotype correlation in a case of moderate clinical
RT severity.";
RN J. Invest. Dermatol. 106:119-124 (1996).
RN [16]
RP VARIANT DEB ARG-1782.
RX MEDLINE=96183562; PubMed=8618018;
RA Christiano A.M., McGrath J.A., Uitto J.;
RT "Influence of the second COL7A1 mutation in determining the
RT phenotypic severity of recessive dystrophic epidermolysis bullosa.";
RN J. Invest. Dermatol. 106:766-770 (1996).
RN [17]
RP VARIANT DEB ASP-2073.
RX MEDLINE=96310789; PubMed=8757758;
RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,
RA Uitto J., Pope F.M., Eady R.A.J.;
RT "Clinicopathological correlations of compound heterozygous COL7A1
RT mutations in recessive dystrophic epidermolysis bullosa.";
RN J. Invest. Dermatol. 107:171-177 (1996).
RN [18]
RP VARIANTS DEB TRP-1982; GLY-2008; ALA-2025; GLU-2049; TRP-2063 AND
RP ARG-2575.
RX MEDLINE=97465605; PubMed=9326325;
RA Hovnanian A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,
RA Fraïtag S., Christiano A.M., Uitto J., Lathrop M., Barrandon Y.,
RA de Prost Y.;
RT "Characterization of 18 new mutations in COL7A1 in recessive
RT dystrophic epidermolysis bullosa provides evidence for distinct
RT molecular mechanisms underlying defective anchoring fibril
RT formation.";
RN Am. J. Hum. Genet. 61:599-610 (1997).
RN [19]
RP VARIANT DEB ARG-1652.
RX MEDLINE=98106792; PubMed=9444387;
RA Cserhalmi-Friedman P.B., Karpatis S., Horvath A., Christiano A.M.;
RT "Identification of a glycine substitution and a splice site mutation
RT in the type VII collagen gene in a proband with mitis recessive
RT dystrophic epidermolysis bullosa.";
RN Arch. Dermatol. Res. 289:640-645 (1997).
RN [20]

RP VARIANTS DEB ARG-2009 AND ARG-2043.
RX MEDLINE=97358588; PubMed=9215684;
RA Winberg J.-O., Hammami-Hausali N., Nilsson O., Anton-Lamprecht I.,
RA Navlor S.L., Kerbacher K., Zimmermann M., Krajci P.,
RA Gedde-Dahl T. Jr., Bruckner-Tuderman L.;
RT "Modulation of disease severity of dystrophic epidermolysis bullosa by
RT a splice site mutation in combination with a missense mutation in the
RT COL7A1 gene.";
RN Hum. Mol. Genet. 6:1125-1135 (1997).
RN [21]
RP VARIANTS DEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.
RX MEDLINE=98334662; PubMed=9668111;
RA Hammami-Hausali N., Schumann H., Raghunath M., Kilgus O., Luethi U.,
RA Luger T., Bruckner-Tuderman L.;
RT "Some, but not all, glycine substitution mutations in COL7A1 result in
RT intracellular accumulation of collagen VII, loss of anchoring
RT fibrils, and skin blistering.";
RN J. Biol. Chem. 273:19228-19234 (1998).
RN [22]
RP VARIANTS DEB CYS-2008; ARG-2207 AND SER-2775.
RX MEDLINE=98410969; PubMed=9740253;
RA Kon A., Pulkkinen L., Ishida-Yamanoto A., Hashimoto I., Uitto J.;
RT "Novel COL7A1 mutations in dystrophic forms of epidermolysis
RT bullosa.";
RN J. Invest. Dermatol. 111:534-537 (1998).
RN [23]
RP VARIANT DEB ARG-1347.
RX MEDLINE=99019477; PubMed=9804332;
RA Terracina M., Posteraro P., Schubert M., Sonogo G., Atzori F.,
Query Match 3.0%; Score 390; DB 1; Length 2944;
Best Local Similarity 22.0%; Pred. No. 1.3e-05;
Matches 434; Conservative 141; Mismatches 773; Indels 624; Gaps 106;
QY 726 ASGNEVPKGECSGPAT--VNNSSDTESIPSPHTEAAKDTGNGPK-----PPATIGADGP 778
Db 1303 ADGRPGSPGRAGNPGTGCAGLKGSGPLGPRGDPG-ERGRPGKPGCAPQVIGEGP 1361
QY 779 -----PPGPPTP-----PR-----RTSRPIETPASEATG 804
Db 1362 GLPKRGDPGSGPPGPRGPLGDPGPRGPGPLGTAMKDGDRGERGPPGEGGIAPG 1421
QY 805 APTPPPPAPSPSAPPVVPV---KEEKEBETAAP-----PVEEGEE---QKPPAABELAVD 854
Db 1422 EFGPLGLPGSPGPGQPGVPGPKGKGEKGDSEDAFGLPQPGSPGQGRPGPGALGPKGD 1481
QY 855 TGKAEPPVKSECTEEAEEGP-----AKGDAABAAEATAEGALKAEKKEGGGRATTAKSS 909
Db 1482 RG-FPGPL-GEAGEKEGPPGPPAGSRGLPGVAGRPGAKG---PEGPPGTGRQCEKGE 1536
QY 910 GAPQSDSSATCSADEVDEAEAGDKNRLLSR-----PSLLTFTGDPANASP- 957
Db 1537 GRP--GDPVAVGPAVAGPKGKGDVGP--AGPRGATGVQGERGPPGLVLP-GDPGPKGDPG 1592
QY 958 -OKPLDLKQLKQRAAAATPIQVTKVHEPPEDDAAPTAPAPPPPPQNLQPDAPQOPG 1016
Db 1593 DRGFIGL-----TGRAGPPG---SGPPGKGDGPRCPGPGP-----1626
QY 1017 SSPRGKSRSPAPPADKEAFAAEAKLPGDPCWTSGLPFPVPPREVIVIKASHPADPSAFS 1076
Db 1627 VGPRGRDGEVGEKGDGEGP-----PGDP-----GLPGKAGERG-LRGAPVGRGVGK 1672
QY 1077 -----YAPPGHPLPLGLHDTAPVLPRPTTINPPPLISSAKHPSVLEIQIGAI 1125
Db 1673 GDQGDGPDGDRNGSPGSSGPKG--DRGEPGPGPP-----GRIVDTGPGAR 1716
QY 1126 SQGMSVOLHVYPYSEHAKAPVGPV-TMGLPLPMDPKLAPFSGVKQQLSP--RGAGPPE 1182
Db 1717 EKG-----EPDRGQEGPRGPKDGLPGAPGERGIEGFRGPPGQDQGVRGAGK 1770
QY 1183 SLGVPTAQEAASVL---RGTALGVSFGSSITK-----GIFSTR-----VPSDSAITYRGS 1228
Db 1771 DRGPPGLDGRSLDGKFGAAGSPGNGAAGKAGDPRDGLPGLRGEQGLPGPSG-----1824

Qy 1229 ITHGTPADVLYKGTITRIIGEDSPSLRDGREDLSLPGKHVITYEGKGHVLSYEGGMSVTO 1288
Db 1925 -PPGLPKP-----GEGKPGKNGKNGEPDGE---DGRK-----EKGDS--G 1863
Qy 1289 CSKEDGRSSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRALIPPERHSPHILKEQ 1348
Db 1864 ASGREGR--DGPKEGRAP-----GILGPGQPPGLPGP----- 1894
Qy 1349 HHIRGSIQTQIPRYVFAQEDYLREAKLLKRGTP-----PPPPSRDL---TEAY 1397
Db 1895 ---VGPPQCGPPVPGGTGPKGDRGETGSKGEQGLGGERGLRGEPGVPVNDRLLETAGI 1951
Qy 1398 KTOALGPLKPAHEGLVATVKEAGRSIHEIPRELRLHTPLAPR--PLKEGSIQTGT 1455
Db 1952 KASAL-----REIVETWDESGSGLFVP--ERRRGPKGDSGEQPPKGEPI--GF 1998
Qy 1456 P---LKYDTG-----ASTTSGKHDVRS--LIGSPGRTFPFVHPLDVMADARALERA 1502
Db 1999 PGERGLAGDRDGPQPPGGLALGERGPPGSPGLAGBPGKPIGCLP-----GRAGGVGEA 2054
Qy 1503 CYESLSKSRPQTASSG-----GSIARGAPVIVPELCKPROSLTYEDHGAPPAHGLPRG 1557
Db 2055 -----GRPGERGERGKGERGQGRDGPGLP--GTP-----GPPG---PPG 2091
Qy 1558 SPVTMRBPTPL--QEG--SLSSSKASQDRKLTSTPREIAKSPHSTVPEHHP-----H 1606
Db 2092 PKVSVDHPGLSGEQGPGKGLGAKGEPGSGNDGPKGDRVPGIKGDRGEFGRGDGN 2151
Qy 1607 PISPYEHLRGVSGVDLYSHRIPLAFDPTSIPIRGIPLDAAAAYLPHRLAPNPT--YPHLY 1665
Db 2152 PGLPGE--RCMAG-----PEGKPLGQRPGRPPGPGVGG-----HGDRGPGCAPGLA 2194
Qy 1666 PPLYIRGVPDPAALENRQTIINDYITSQOMHNTATAMAQADMLRGLSPRESSIALNYA 1725
Db 2195 GP--AGPQSPGLKGE-----GETGPPGRGLTGTGAVGLPGP 2231
Qy 1726 AGPRGIIDLSQVPHLVLVPPT--PGTP---ATAMDRALVLPAPQFSSRHSPLSP 1779
Db 2232 PPGSLGVPGQSGPLPGVGTGTGPGAGRGAGKGDGDRSGVP-----GSGPLP 2293
Qy 1780 G--GPTHLTKTPTTSSSRERDRDREREKSIILTSTTVEHAPIWPMPQEQSGS 1837
Db 2284 GPVPGKBPPTGAPG-----QAVVGLPG--AKGE 2311
Qy 1838 SGSSGG--GGSSSRPASHAHQHSPISTQDALQORPSVLHNTGMKIITAVEPSKPT 1896
Db 2312 KGAPGLAGDLVGEFGAKGDRGLPGRGKEGACRAGEPGDPGEDQKG-----APGPKG 2366
Qy 1897 VLRSTSTSSPVKPAATFPATHCPGLGTLGQVYPTLMBPVLLPKEAPRVARPER--PRAD 1954
Db 2367 PKGDPGVGVSPGPPGPPGVKGLDGL--LPG-----LPGAPGVVPGQTGRGE 2414
Qy 1955 -----TGHAFLAKPARSGLEPASSPSKSEPRPLVPVPSGHATIAKTNLAPHAS 2008
Db 2415 MGQPGSGERGLAGPPGREGI-----PGPLGPP-----GPPGSGVPPGAS 2454
Qy 2009 ---PPP-----PAPPASADPHREKTQSKPFSIQELESRLGSLVHGSYSYSGEV--EPV 2056
Db 2455 GLKGDGDPGVGLPGRGERGEPGIRGEDRP-----GOEG-----PRGLTGPPG 2499
Qy 2057 S-----PVSSPSBLTHDK-----GLPKHLELDKSHLEGELRPK-----QP 2091
Db 2500 SRGERGKDVSGAKLGDGSDASVILPGPPGRGAKGDMGREGPLGDGKPRGNDGP 2559
Qy 2092 GPVKLGGE-----AAHLPHLRPL--PESQPS-----SPLIQTAPGVKHQ----- 2130
Db 2560 GDKSGKEPDKGSAGLPGLRGLLGPGQPGAAIPGPGSPGKDVPGIRGEKGDVGFM 2619
Qy 2131 --RVVTLAQHISEVITQDTRHHPPQOLSAPLPAPLYSPFGASCPVLRLRRPPSDLYLPPP 2188
Db 2620 GPRGLKGERGVKACGLDGEKDGKGEAGPP-----GRPGLAGHKGEMGEFG-----VPG 2668

Qy 2189 DHGAPAR-----GSPHSEGGKSPFNKTSVLGG-----GEDG-IPVSP-- 2227
Db 2669 QSGAPGKEGLIGPKGDRGDFGQPCPKGDQCEKGERGTPGIGFPFGPSGNDGSAGPPGPG 2728
Qy 2228 -----PEGMT-----EPGHSRAVYPLLYRDGEQTEPSRMGSKSPGNTSQPPAFSK 2274
Db 2729 SVGPGRGEGLOQKGERGPPGERVVGAPGVGAPGERGEQGRPGFAGRGEKGEAA----- 2784
Qy 2275 LTESNSAMVSKKQEKINKLNTN-----NRNEPEYNISQPGTEIFNMPAITGTG 2323
Db 2785 LTEDDI-----RGFVQEMSOHCACQOFIASSGRPLPSYAADTAGSQLHAVFLR--- 2835
Qy 2324 LMTYRQAOVQEHASTNMGLEAIRKALMGKYDQWESPPISANAFNLNA--SASLPAAM 2381
Db 2836 ---VSHAESE-----ERVPPEDDEYSEYSEYVQDPEAPWDSDDDPCSL 2878
Qy 2382 PI-----TAADGRSD--HTLTSPPGGGKAKVSGRPSRKAQSP 2417
Db 2879 PLDEGSCCTAYTLRWYHRAVTGSTEACHPFVYVGGCGGNANRGTREACERRCP 2930

RESULT 36
HRX_MOUSE
ID HRX_MOUSE STANDARD; PRT; 3866 AA.
AC P55200;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Zinc finger protein HRX (ALL-1) (Fragment).
GN MLL OR HRX OR ALL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=C57BL/6J, and C57BL/6 X CBA; TISSUE=Lung, and Spleen;
RA MEDLINE=933117679; PubMed=8327517;
RX Ma Q., Alder H., Nelson K.K., Chatterjee D., Gu Y., Nakamura T.,
RA Canaan E., Croce C.M., Siracusa L.D., Buchberg A.M.;
RT "Analysis of the murine All-1 gene reveals conserved domains with
RT human ALL-1 and identifies a motif shared with DNA
RT methyltransferases."
RL Proc. Natl. Acad. Sci. U.S.A. 90:6350-6354 (1993).
CC -!- FUNCTION: Possibly acts as a transcriptional regulatory factor.
CC May regulate genes involved in skeletal formation during embryonic
CC development.
CC -!- SUBUNIT: Interacts with SBF1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P55200-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P55200-2; Sequence=VSP 006667;
CC -!- SIMILARITY: Belongs to the TRX/MLL family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 SET domain.
CC -!- SIMILARITY: Contains 3 PHD-type zinc fingers.
CC -!- SIMILARITY: Contains 1 CXXC-type zinc finger.
CC -!- SIMILARITY: Contains 1 post-SET domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L17069; AAA62593.1; -;
CC MGD; MGI:96995; MLL.
CC GO; GO:0005634; C:nucleus; IDA.
DR
DR

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;
RP GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND
RP CYS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;
RP PRO-574; GLU-1269 AND PRO-1474.
RX MEDLINE=21064696; PubMed=11134255;
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT "Structure of the human type IV collagen gene COL4A3 and mutations in
RT autosomal Alport syndrome.";
RL J. Am. Soc. Nephrol. 12:97-106(2001).
RN [4]
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93015826; PubMed=1400291;
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
RT Goodpasture antigen (alpha 3(IV)NCL). Identification of a potentially
RT antigenic region at the triple helix/NCL domain junction.";
RL J. Biol. Chem. 267:19780-19784(1992).
RN [5]
RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE=91353570; PubMed=1982840;
RA Morrison K.E., Mariyama M., Yang-Peng T.L., Readers S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RL Am. J. Hum. Genet. 49:545-554(1991).
RN [6]
RP SEQUENCE OF 1331-1670 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92147878; PubMed=1737849;
RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,
RA Pusey C.D.;
RT "Molecular cloning of the human Goodpasture antigen demonstrates it
RT to be the alpha 3 chain of type IV collagen.";
RL J. Clin. Invest. 89:592-601(1992).
RN [7]
RP SEQUENCE OF 1644-1670 FROM N.A.
RC TISSUE=Kidney;
RA Ding J.;
RN Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RC TISSUE=Kidney;
RX MEDLINE=94124597; PubMed=8294492;
RA Feng L., Xia Y., Wilson C.B.;
RT "Alternative splicing of the NCL domain of the human alpha 3(IV)
RT collagen gene. Differential expression of mRNA transcripts that
RT predict three protein variants with distinct carboxyl regions.";
RL J. Biol. Chem. 269:2342-2348(1994).
RN [9]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=98196854; PubMed=9537506;
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Ninomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
RT 2q36.";
RL FEBS Lett. 424:11-16(1998).
RN [10]
RP ALTERNATIVE SPLICING.
RX MEDLINE=93280184; PubMed=8505332;
RA Bernal D., Quinones S., Saus J.;
RT "The human mRNA encoding the Goodpasture antigen is alternatively
RT spliced.";
RL J. Biol. Chem. 268:12090-12094(1993).
RN [11]
RP VARIANT PRO-1474.
RX MEDLINE=95078827; PubMed=7987301;
RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
RA Readers S.T., Smets H.J.M.;
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal

recessive Alport syndrome.";
Hum. Mol. Genet. 3:1269-1273(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist. Isoforms differ in
CC the C-terminal part of the NCL domain;
CC Name=1;
CC IsoId=Q01955-1; Sequence=Displayed;
CC Name=2; Synonyms=V;
CC IsoId=Q01955-2; Sequence=VSP_001170;
CC Name=3; Synonyms=LS;
CC IsoId=Q01955-3; Sequence=VSP_001171;
CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCL) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Isoform 2 contains an additional N-linked glycosylation site.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCL domain, are conserved in all known type
CC IV collagens.
CC -!- PTM: Phosphorylated by the Goodpasture antigen-binding protein.
CC -!- DISEASE: Antibodies against the NCL domain of alpha3(IV) mediate
CC the autoimmune disease Goodpasture syndrome (MIM:233450), which is
CC characterized by hematuria and pulmonary hemorrhage.
CC -!- DISEASE: Defects in COL4A3 are a cause of autosomal recessive
CC Alport syndrome (AS) (MIM:203780), an hereditary disorder
CC characterized by progressive glomerulonephritis, renal failure,
CC hematuria, ocular abnormalities and deafness. The recessive form
CC occurs equally between males and females.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X80031; CAA56335.1; -;
DR EMBL; AJ288487; CAC36101.1; -;
DR EMBL; AJ288488; CAC36101.1; JOINED.
DR EMBL; AJ288489; CAC36101.1; JOINED.
DR EMBL; AJ288490; CAC36101.1; JOINED.
DR EMBL; AJ288491; CAC36101.1; JOINED.
DR EMBL; AJ288492; CAC36101.1; JOINED.
DR EMBL; AJ288493; CAC36101.1; JOINED.
DR EMBL; AJ288494; CAC36101.1; JOINED.
DR EMBL; AJ288495; CAC36101.1; JOINED.
DR EMBL; AJ288496; CAC36101.1; JOINED.
DR EMBL; AJ288497; CAC36101.1; JOINED.
DR EMBL; AJ288498; CAC36101.1; JOINED.
DR EMBL; AJ288499; CAC36101.1; JOINED.
DR EMBL; AJ288500; CAC36101.1; JOINED.
DR EMBL; AJ288501; CAC36101.1; JOINED.
DR EMBL; AJ288502; CAC36101.1; JOINED.
DR EMBL; AJ288503; CAC36101.1; JOINED.
DR EMBL; AJ288504; CAC36101.1; JOINED.
DR EMBL; AJ288505; CAC36101.1; JOINED.

Qy	654	FNFKKRNLDLILQOHLKMEKERNARRKKQAPAAASEEAFPPVVEDEEMEASGVSGN	713
Db	1189	CKMRKCNLQ--WMPSKAYLQKAKAVKKEKSKTS-----EKDKSESSVVKV	1236
Qy	714	EEEMVBEAEALHASGNVPRGCSG-----PATVNNSSDTSISIPSPH-----TAAKD	761
Db	1237	---VVDSSQKPTSAERDPAPKXSSSPPPRKPVVEEKSEBGNVSAPGPESKQATTASRK	1293
Qy	762	TGQNGPKPPATLGNAGPPPGPT--PPRRTSRAPITPASEATGATPPPPAPSPS---	816
Db	1294	SSQVQSOPALVI-----PPQPTTGPBRK--EVP-KTTPSEPKKQPPPPSPESGPESQK	1345
Qy	817	--APPVVVPEKEEETAAAPVEEGS-----EOKPPA-----	847
Db	1346	KVAPRPSIPVKQPKPEKEKPPVNVKQENAGTILNLSTLSNGNSKKQIPADGVHRTVDP	1405
Qy	848	---ABE-----LAVDTGAE-----EPVKSECTBEAE	872
Db	1406	KEDCEAVNWMGGLILTSVPIPTRVVCFLCASSGHVEPVYQVCEPFPKFCLEB-NE	1464
Qy	873	GPAK-----GKDAEAAEATAE-----GALLAEKE--	897
Db	1465	RPLEDQLENCCRRCKFCHVCGRQHQATKOLLECNKCRNSYHPCELGPNYPTKTKKKV	1524
Qy	898	-----GSGRATTAKSSGAPQDSGSA-----TCSADEVEAE	930
Db	1525	WICTKVCVKSCSGTTPGKQWDQASHDFSLCHDCAKFAKGNFCLPCLDKYDDDDYESK	1584
Qy	931	-----GGDKNLLSPRSLTPT-----GDPANASQKPLDL	963
Db	1585	MMQCGKCDRWVHSHKCNLSDEMYEILLNPESVAYTCVNCTERHPAEWRLALEKEQLISL	1644
Qy	964	KQ-----LKQRAAALPPQVTKVHEPPREDAAATPKPAPPAAPPPQONLOPES	1009
Db	1645	KQVLTALLNGRTTSHLLRYQAAKPPDL-----NPETEESIPSRSPGEGDPPLVTEVSK	1699
Qy	1010	DAPOQCGSSPRGKSRSPAPPADKEAFAAEAKLPGDPCCWTGSLPPVPVPREVIKASPA	1069
Db	1700	QDDQP-----LDLEGVKRMDGNTITSLEFSDDDIVKLIQAINSL	1740
Qy	1070	PDFSAFSYAPPGLPLGLHDTARP-----VLP-----RPPTISNPPLISS	1111
Db	1741	DG-----GQPEIKKANSWKSPFIRQWERMVFPWFVSKSRFNEPNKVSNSGMLPN	1791
Qy	1112	AKHPSVL-----ERQIGA-ISQMSVOLHPYSEHAKAPGVPTMGLPLPMDPKCLA	1162
Db	1792	AVLPPLSDHNYAQWQEREENSHTEQPLMKKIIP---APKPGKGPDSPTPLHPPTPP	1847
Qy	1163	PFGVKQEQLSPRQAGPPESL-----GVPTAOEA-----	1192
Db	1848	ILSTDRSREDSP--ELNPPPGIEDNRICALCTVGDSDSADAGRLLYIGQNEWTHVNCAL	1905
Qy	1193	-----SVLRGTALG-----SVPGGSIKGIPISTRVPFSDSAITRGSIIT	1230
Db	1906	WSAEVFEDDDGSLKNVHMAVIRKQLRCEFCQEGATV--GCCLTSCTSN-----	1953
Qy	1231	HGTPADVLYKGTITR-----IGEDSPSRLDRGREDSLPKGHVI-----YEG	1272
Db	1954	-----YHFMCSRANKCVFLDDKKVYQQRHRD--LIKGEVVPENGFEVFRVFDVFE	2003
Qy	1273	-----KKGHVLSYEGMSVT-----CQSKEDGRSSGPP	1301
Db	2004	ISLRKFLNGLEPENIHMMIGSMITDCLGILNDLSDCEDKLFPIGQCS-----	2052
Qy	1302	HETAAPKRTYDMBGRVGRATISSASIEGLMGRATIPPRHSPHFLKEQHIRGSIITQGI	1361
Db	2053	-----RVVWSITTDARKRCVYTKVIECRPPVVEPDINSTEVDENHTAHS-----	2100
Qy	1362	SYVEAQEDYLREAKLAKRGTEGTPPPPPSRDLTEAY-----KTQALGPLKUKPAH	1411
Db	2101	SFTBSSKESONTAEIIS-PPSPDRPHSQTSGSCYHVHISKVPRINTSPYSTORSPOG	2159

Qy	1412	EGLVATVKAGRSIH	PREELRHTPELPA	RPLKEGSI	TGOTPLKYDTGASTT	GSKKH	1471	
Db	2160	RPL	PSAG	SPTPTTH	IVTVGDPL	LSSGLRSIGSRH	2195	
Qy	1472	DVPSLIGSCRTPP	VHPLDVADAPAL	ERACVEB	SLKSRPGTASS	SGSITARGAPVIV	1530	
Db	2196	STSSL	SPQRS	KLRIMS	PMKTGNTY	GRNNVSSVTGTAT	2244	
Qy	1531	PELGKPRQSP	LYEDHCAPFA	GHLPRGS	PVTWREPT	PRLQEGSLSS	1577	
Db	2245	HVLG	PLNAST	LGQNTST	SNLQRTVTVGN	KNSHLDGSSSEMKQSSASDLVKSSSL	2303	
Qy	1578	KASQDRKLSTP	REIAKSPHSTV	PEHHPHI	PGYEHLLRGV	GVDLRSHIPLAFDP	1637	
Db	2304	KGEKTKVLSK	SSSE	GSAAHN	VAYPGIP	KLAPQVH	2357	
Qy	1638	PRGIPDAAA	AYLPHLAP	NTYPHLY	PPYLIRGY	PD	1675	
Db	2358	SFSSK	REAL	SFPHLH	LRQQRNDRDO	HTDSTQSANSP	2399	
Qy	1676	TAALENRQ	TIINDYI	TSQOMEHN			1698	
Db	2400	KTLKLSGMS	NRSSI	NEHMGSS	RRRQK	KKSKETFEKHSSK	2459	
Qy	1699	TATAMA	QR	ADML	RGLS	PRESS	1719	
Db	2460	LKPEFDEV	LTPYMG	QRPCNN	VSSDKIGK	GLSMPGVKAPPMQ	VEGSAKELQAPRKT	2519
Qy	1720	LAL	NYAAG	PRGI	IDLQVPHL	PVLVPPT	PGTPATAM	2577
Db	2520	VKVTL	TPLK	MENESQ	KNALKESS	PASPLQ	TESTSPTEPIS	2577
Qy	1768	FSSRH	SSPLS	PGGP	THLTP	TTTSSSER	ERDRDRDREREK	1815
Db	2578	TSQDSQ	SQNNYQ	NLPQDR	NLMLDGP	KPEDGS	FKRYP	2628
Qy	1816	TSTTV	EHA	IRW	PGTE	QSSGSSG		1843
Db	2629	FGLT	PLYG	VR	SYGE	DPFYSS	TGKKR	2688
Qy	1844	GGSSSR	PASHS					1855
Db	2689	ISSG	GERLASH	NLFRE	EQCDLPK	I	SOLD	2748
Qy	1856							1882
Db	2749	GTE	NLKID	RED	AGEKH	VTSSVGH	KNEPKM	2807
Qy	1883	MKGII	TA	VEP	SKEPT			1906
Db	2808	RVHT	TS	DKN	LLD	TYNTELLK	SDSDNN	2862
Qy	1907	V	RPA	TP	PPATH	CP	LG	1936
Db	2863	LG	SPSS	SEL	NL	EG	GLD	2921
Qy	1937	LL	KEA	PR	VAR	PER	PRAD	1977
Db	2922	EL	LE	UP				2967
Qy	1978	GS	PR	PL	VP	VS	G	2023
Db	2968	AS	ESD	P	ALLS	P	G	3017
Qy	2024	E	K	T	O	S	K	2082
Db	3018	NN	QD	TRNS				3052
Qy	2083	E	G	E	L	R	P	2133
Db	3053	SP						

Db 3106 SVSSTPVMETNTSVLGMGGLTLTGLNPSLTSOSLFPSSKGLLPMSHHQHLHSF 3165
QY 2175 -----DLRRPPDLXL---PPDHGAPGSPH-----SEGGKSP--- 2207
Db 3166 PAATQSSPPPNISNPPSGLLIGVQPPDPOLLVSSSQRTDLSTVATPSSGLKKRPISR 3225
QY 2208 ---EPNKTSLVGGGDBGIEPVSPPGMT-----EPQH-----SRSAYVPLL 2245
Db 3226 LQTRKNKLLAPSGTSPNIAPSDVWSNMTLINFTPQLNHPSPLLDLGSLNTSSHRTVENI 3285
QY 2246 YRDGQT-----EPRMGSKSPGNTQPPAPFSPKLTESAMVK-----SKQELNKKLN 2295
Db 3286 IKRSKSSIMYFEPAPLFPQSGVGTATAAGTSTISQDTSHLTSGSVSGLASSSSVLNVVS 3345
QY 2296 THNRNEPEVNIISOPGTEIFNMPAITGTLMTYRSQAVOEHAHNMGLEAIRKALMGKYD 2355
Db 3346 MOTTTPPTSSASVPHGVHVTNPRLLGTDPDGISNLLIKASQOSLGIO-----D 3394
QY 2356 QWEEFPLLSANAFNPLNASALPAAAMPITAADG-----RSDHTLSPGGGK----- 2402
Db 3395 QPVALPP-SSGMFPQLGTSQT-PSTAATAASSICVLPSTQTTGITAASPSGEADHYQL 3452
QY 2403 -----AKVSGRPSRKAQSPAGLSDRPPSVSVHSEGCNRRTPLNVRVWEDRPS 2455
Db 3453 QHVNQLLASKTGIHSSQRDLDSASG-----POVSN-----FTQTV--DAPN 3491
QY 2456 SAGSTPPFPYPLIMLQAGVMAAPPPGLPAGSGPLAGHHAWDEPKP 2504
Db 3492 SMG-----LEQN-----KALSSAVQASPTSPG-GSPSPSSGQRSASPSVPGP 3533
RESULT 40
CA54_HUMAN
ID CA54_HUMAN STANDARD; PRT; 1685 AA.
AC P29400; Q16006; Q16126;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 5 (IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614 (1994).
RN [2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC TISSUE=Kidney;
RX MEDLINE=923116923; PubMed=1352287;
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481 (1992).
RN [3]
RP SEQUENCE OF 85-1685 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=90337990; PubMed=2380186;
RA Pihlajaniemi T., Pohjola E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5 (IV).";
RL J. Biol. Chem. 265:13758-13766 (1990).
RN [4]
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE=91169491; PubMed=2004755;

RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9 (1991).
RN [5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610 (1990).
RN [6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjola E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5 (IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";
RL Am. J. Hum. Genet. 46:1024-1033 (1990).
RN [7]
RP SEQUENCE OF 1-20 FROM N.A.
RX Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBAJ databases.
RN [8]
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NCI domain.";
RL Kidney Int. 44:1316-1321 (1993).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499 (1997).
RN [10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;
RT "Single base mutation in alpha 5 (IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18 (1991).
RN [11]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142 (1992).
RN [12]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129 (1992).
RN [13]
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggama-Schouten W.A.G., Roodvoets A.P., Rascher W.,

RA van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of
RT patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
[14]
RP VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-796;
RP ARG-869; ARG-872 AND CYS-1241.
RX MEDLINE=95322976; PubMed=7599631;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RA "Detection of 12 novel mutations in the collagenous domain of the
RT COL4A5 gene in Alport syndrome patients.";
RL Hum. Mutat. 5:197-204(1995).
[15]
RP VARIANTS AS ARG-1649.
RX MEDLINE=96213750; PubMed=8651292;
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
RA Denison J.C., Fain P.R., Gregory M.C.;
RT "A mutation causing Alport syndrome with tardive hearing loss is
RT common in the western United States.";
RL Am. J. Hum. Genet. 58:1157-1165(1996).
[16]
RP VARIANTS AS.
RX MEDLINE=96213754; PubMed=8651296;
RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
RA Savi M., Ballabio A., de Marchi M.;
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
RT exons of the COL4A5 gene.";
RL Am. J. Hum. Genet. 58:1192-1204(1996).
[17]
RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
RP MET-1428.
RX MEDLINE=97094179; PubMed=8940267;
RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,
RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
RA Gubler M.-C., Antignac C.;
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
RT syndrome.";
RL Am. J. Hum. Genet. 59:1221-1232(1996).
[18]
RP VARIANTS AS ASP-1498.
RX MEDLINE=96233932; PubMed=8829632;
RA Tverskaya S., Bobryina V., Tsalykova F., Ignatova M.,
RA Krasnopolekaya X., Evgrafov O.;
RT "Substitution of A1498D in noncollagen domain of $\alpha 5(\text{IV})$ collagen
RT chain associated with adult-onset X-linked Alport syndrome.";
RL Hum. Mutat. 7:149-150(1996).
[19]
RP VARIANTS AS GLN-1677.
RX MEDLINE=97295089; PubMed=9150741;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT "Common ancestry of three Ashkenazi-American families with Alport
RT syndrome and COL4A5 R1677Q.";
RL Hum. Genet. 99:681-684(1997).
[20]
RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
RP AND ASP-1596.
RX MEDLINE=98112435; PubMed=9452056;
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
RA Trivelli A., Pinciattoli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;
RT "Missense mutations in the COL4A5 gene in patients with X-linked
RT Alport syndrome.";
RL Hum. Mutat. Suppl. 1:S106-S109(1998).
[21]
RP VARIANTS AS VAL-420; 456-PRO--PRO-458 DEL; ASP-573; ASP-624; ASP-635;
RP 802-GLY--PRO-807 DEL; ARG-869; CYS-941; SER-941; SER-1030; SER-1066; ASP-1143;
RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.
RX MEDLINE=99063529; PubMed=9848783;
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
RA Barker D.F., Gregory M.C., Atkin C.L., Stykardottir U., Neumann H.,
RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;

RT "High mutation detection rate in the COL4A5 collagen gene in suspected
RT Alport syndrome using PCR and direct DNA sequencing.";
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
[22]
RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
RX MEDLINE=20030197; PubMed=10561141;
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
RT "Detection of mutations in the COL4A5 gene in over 90% of male
RT patients with X-linked Alport's syndrome by RT-PCR and direct
RT sequencing.";
RL Am. J. Kidney Dis. 34:854-862(1999).
[23]
RP VARIANTS AS ARG-822.
Query Match 2.9%; Score 379; DB 1; Length 1685;
Best Local Similarity 20.7%; Pred. No. 1.8e-05;
Matches 380; Conservative 136; Mismatches 638; Indels 680; Gaps 90;
QY 764 QNGPKPEATLGDGPP-----PGPPTPRRTSRAPIBPTPASEATGATPTPPAPP-- 813
Db 143 QGPPGPGIOMKEGPGSIISSLPFGKGNPGVPGPGIQLPGPTGIPGIGPGPGPL 202
QY 814 -SPSAPPVV-----PKEEKEETAAPVVEGE--EQKPPAAAE-----L 851
Db 203 MGPPGPGLPKPKNMGLNFQKGEQGLQGPFGPGQISQKRPIDVFQKQDGL 262
QY 852 AVDTGKAEPVKSCTEAEAEPAKGDAAEAETAAGALKAEKKEGSGRAITAKSSGA 911
Db 263 PGDRGPPGPP-----GIRGPPGPGKGEQFGKRGKPGKGDGNGQPGI---PGL 314
QY 912 PQDSSSATCSADVEDAEAGDKN-----RLLSRP----- 942
Db 315 PGDEYEGEPGRDG-EKGQKGDTPGPPGPGVLPRTGTITGEKNIGLPGLPGEKER 373
QY 943 -----SLLTPTGDP-----RANASPKPLDLK-----QLKORAAAI 973
Db 374 GFPGIQGPPGLPGPPGAAMVPGPPGPPGPPGGERGQKDEGPPGISIPGPGLDGQGPAPGL 433
QY 974 P-PIQVTKVHEPPRED-AATKPAAPPAPPPONLOPE-----SDAP 1012
Db 434 PGPEGPAAGHIPPSDEICEPFGPPGPGSGDKLQGEQGVKGDGTCFCNIGTIGSGPP 493
QY 1013 QQQP-----SSPRGKSRSAPPADK-EAFAAAEAOPLGDPDPCWTSGLP-FP---VPPREVI 1063
Db 494 GQPLGPGPGSLGPGQKGEQAGATGPKGLGIP--GAPGAPGFGSKGEPGDIL 551
QY 1064 K-----ASPHAP-----DPSAFSY-----APFGHPLPL 1086
Db 552 TFPCKMGDKGELSGFAPGLPGTPGQDGLPGKPGKBPGGITFKGERGPPGNPGLP 611
QY 1087 GLHTAPVLPRTPTISNPPLISSAKHPSVLERQICAIQSGMSVOLHPVSEHAKAPVG 1146
Db 612 GLPGNIGPM--GPPGFGPPG-----VGEKGIQGA-GNPGQGP- 653
QY 1147 -----PVTMGLP-----LPMDPKLAPFSGVKQEQLSRPGQAGPESLGVPT 1188
Db 654 DPGQTIQPKPGPLPGNPRDGDVGLGDDP-----GLPQPGPLGPG 696
QY 1189 AQEASVLRTALGALGVPVGGSTIKGIPSTRVPDSAITRGSITHTGTPADVLYKGTITRIIG 1248
Db 697 SKGEPGPIGLPGPPG--PKGFFGIPGPPGAP-----GTPGRIGLEG----- 737
QY 1249 EDSRSLDRGREDSLPKGHVIECKKHVLSYEGGMSVTCQSKEDGRSSSSPPHETAAPK 1308
Db 738 -----PPGPGFPGPKG-----BPGFALPGFPGPGPL- 765
QY 1309 RTYDMEGRVGRAISSASIEGLMGRAPIPPERHSPHLKEQHHRGSIQTGIPRSYVEAQE 1368
Db 766 -----GFKG-ALGPKGDRGFPPGPP----- 786
QY 1369 DYLRREAKLKRGTGTPPPPPSRDLTBAYKTQALGPLKL-----KPAHEGLVATVKEA 1421

Db 787 -----GRTGLDGLPGKGDVGNPGPGLPGVGVQPPGPGIPIGPQOP 836
Qy 1422 GRSIHEIPRE-----LRHTPLPLAPRL-KEGSIITOTPLKYDTGAST 1465
Db 837 G--LHGIPGKDPGPGGLVPGPGGSGPGIIPGAPGIPGPGS--PGLPGK--AGASG 890
Qy 1466 TGSKKHVDVRSIGSGRTFPVHPVLDVMDARALACRYESLSRPGTASSSGSIARG 1525
Db 891 PPGTKGEM-GMWGPG-----PFGPLGIPGRS-GVPLGKDDGLQOQPGLPPTGKSGK 944
Qy 1526 APVIVPELGKRSQSLTYEDHAGPAGHLPGSPVTRPEPTPRLOEGLSSSKASQDRKL 1585
Db 945 EP-----GLP-----GPGPM-----DPNLLGSKGEGK-- 967
Qy 1586 TSTPREIAKSPHSTVPEHHHPISPYEHLRGVGVLDLYRSHIPLAFDPTSIPIRGIPLDA 1645
Db 968 -----EPGLPGIPGVSGPKGYQGLPGDP-----GQP-GL 995
Qy 1646 AAAYLPHRLAP--NPTYPH-----LYPPYL-----IRGYPDTAALENRQTI 1685
Db 996 SQQPLGPPGPKGNPGLPGQGLIGPGLKGTIGDMGFPQGVGPGPGSGVGPQ-- 1053
Qy 1686 INDVITSQMHNTATAMAOQADMLRGLSPRESSALNYAAGPGLIIDLQVPHLPVL-- 1743
Db 1054 -----GSPGLPGQKGD--KG-DPGTSSIGLPGLPKPGKGPGLPGVGNPGIKG 1098
Qy 1744 -----VPPTPGTP-ATAMDRLAYLPTAPQFPSSRHSPLSPGQPTHLTXTPTTSSSER 1796
Db 1099 SVGPGPLGPGTPOAKGQGLPGPGT-----GPPGPKGISG-- 1138
Qy 1797 ERDRDRDRDREREKSLTSTTTTVEHAPIWRPGTEQSSGSGSGSGSGSSSRPASHSH 1856
Db 1139 -----PGNPLPGEPGPGVGGG-----H 1156
Qy 1857 AHQSPISPRQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSSVPRPATPPPA 1916
Db 1157 PGQPGPGPKGPGQDGIPG---PAGQKG-----EPGQP-----GFGNPGPG--LPL 1200
Qy 1917 THCPGLGGTLDGV--YPTLMEVLLP--KEAPRVARPERPRADTGHAFIAKPARSGLEPA 1972
Db 1201 SQQKGDGLPGIPGNPGLPGKPGEGFHGPGVQVQPGPGSPGA--LEGPKNGP--PQ 1257
Qy 1973 SSPSKGSPEPLVPP-VSGHATIAPTAKNLAPHASDPDPAPPASAD---PHREKTQS 2028
Db 1258 GPPGRPGLPGEPGPGPLPGNGGI-----KGEKGNPGQPGPLPGLKGLKGDQPGPGLQGNPG 1312
Qy 2029 KPFSIQELELRSLGVHSGSYSP--EGVEPVSPVSPSLTHDKGLPKHLELDKSHLEGE- 2085
Db 1313 RP-----GLNGMKGDPLGPGVPGFPGMKGPS-----GVP-----GSAGPEGE 1350
Qy 2086 --LRPKQPGPVKLGEAAHLPHLRPLPESQPSSSLQTPAGVKGHQRVVTLAQHISEVI 2143
Db 1351 GLIGP--PGPGLPG-----PSGSIITKGAGPGPGIPGQGLK----- 1388
Qy 2144 TDYTRHHPOOLSAPLAPL--YSPPGASCPLVLDLRPPPSLDYLPDPHAGAPARGSPHSE 2201
Db 1389 -----LPQPGQPGLPGLPTGPGD-----PGRNGLPGFDGAGRGDGLP 1429
Qy 2202 GKKSPENKTSVLG--GGEDGIEPVSPPEGTEPGHRSRVSAPVLLYRDBGOTEPFRMGSK 2260
Db 1430 G-----QRTGLDGPDPGLQPGPGPGTSSVAHGF-----LITRHSQTTD---APQ 1475
Qy 2261 SPGNTSOPPAFESKLTESNAWVSKKQ-----EINKKLT 2296
Db 1476 CPOGTQVYEGFSL-----YVQGNKRAHQDGLTAGSCLRRFSTMPFMCNINNVCF 1529
Qy 2297 HNRNEPYNISQPGTEIFNMPAICTGLMYRSQ 2330
Db 1530 ASRNDYSYWLSTPFPMPMSMQPLKQSIQPFISR 1563

RESULT 41

DRPL RAT
ID DRPL RAT STANDARD; PRT; 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Atrophin-1 (Dentatorubral-pallidolysian atrophy protein).
GN DRPLA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.; Young W.S., Li S.-H., Schilling G.,
RT "Cloning and expression of the rat atrophin-I (DRPLA disease gene)
RL homologue.";
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Epplen J.T., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
dentatorubral-pallidolysian atrophy (DRPLA) in rat.";
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -!- SUBUNIT: Interacts with WWP1 and WWP2 (By similarity).
CC -!- TISSUE SPECIFICITY: Predominant neuronal expression, although
markedly reduced amounts are found in most other tissues.
CC -!- DEVELOPMENTAL STAGE: Similar expression at all development stages
(14.5 dpc, 17.5 dpc, newborns and adults).
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U31777; AAA80337.1; -.
CC EMBL; X89453; CAA61623.1; -.
CC InterPro: IPR002951; Atrophin.
CC Pfam; PF03154; Atrophin-1; 2.
CC PRINTS; PR01222; ATROPHIN.
CC DOMAIN 165 171 POLY-PRO.
FT DOMAIN 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-PRO.
FT DOMAIN 477 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT DOMAIN 502 505 POLY-PRO.
FT DOMAIN 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT CONFLICT 455 455 N -> S (IN REF. 2).
FT CONFLICT 594 594 F -> L (IN REF. 2).
FT CONFLICT 689 689 P -> R (IN REF. 2).
FT CONFLICT 717 717 T -> M (IN REF. 2).
FT CONFLICT 737 737 A -> V (IN REF. 2).
FT CONFLICT 965 965 MISSING (IN REF. 2).
SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9B1F CRC64;

Query Match 2.9%; Score 378; DB 1; Length 1183;
Best Local Similarity 20.6%; Pred. No. 1.4e-05;
Matches 321; Conservative 146; Mismatches 604; Indels 486; Gaps 72;
Qy 762 TGQNGPXPATLADGADGPPGPPPTPRRTSRAPIETPTASEATGATPPPPAPSPSPAPPV 821
Db 3 TRQKDSMSMRSGRKEAPGPRELRSRGRASPGCVSTSSDG----- 45

QY 822 VPKEEKEETAAPPVEGEBEQPPAAEEELAVDTGKAEFPVKSECTEEAEEGPAGKDAE 881
Db 46 --KAEKSRQAKKARVEETSTPK-----ANKQGRSEEEISEE--SEETSAKKYKTEE 94
QY 882 AAEATGALKAEKKEGGSRATTAKSGAPOSDDSATCSADEVDEAEGDKNRLLSPR 941
Db 95 LPRQPSDL-----DSLDRSINDGSSDPRDIDQ-----NR--STS 131
QY 942 PSLTPTGDPANASPKPLDLKOLKORAAAIPIQVTKVHEPPREDAAATKAPAPPPP 1001
Db 132 PSIYSP-----GSEVNDSDSSGLSGQARP-----YHPPPL-----FPPSPPP 170
QY 1002 PONI--QPEDAPOQSSPRGKRSAPPADKAEFAAEQAQKPGDPCWTSGLPF--- 1055
Db 171 PDSIPROGSEFEPHPSVPTG-YHAPMEPTSLFOGPP--PGAPPHPQLPYGSGAG 226
QY 1056 -----PVPPREVIKASPHADPSAFSYAPPGLPLGLHDTA-----R 1093
Db 227 GVLGPPMPGKGAASVGPSSGKQHPPTTPIPISSGASGAPPAKPNTPVAGNL 286
QY 1094 PVLPRPTISN-----PPLISSAKHPSVLEROIGALSQMSVOLHVPYSEHAKA----- 1143
Db 287 PSAPPATFPHVTENLPPPPALRLNANASPPGKGA-----QIPGHLPL-SPHAMGQMS 341
QY 1144 --PVGPTWGLPLPMDPKLAPFGVKGQEQSLPRGQAGPPESLGVPTAQASVLRGTALG 1201
Db 342 GLPPGP-EKGPTLAPSPHPLPASS-----SAPGPPMRYPYSSCSSSV----- 384
QY 1202 SVPGSITKGPSTRVPSDSAITYRGSIHTGTADVLVYKGTITRIIGEDSPSLDRGRED 1261
Db 385 ---AASSSSSAATQYPASQTLPL--SYPHSFPPL-----TSMVSNQPKY--TQP 428
QY 1262 SLPKGHVIEYEGK-----GHVL-----SYEGMSVTQCKEDGRSSGPPHETAPKRTY 1311
Db 429 SLPSQAVMSQGPPIPPPPYGRLLPNNTHPGFFPT-----GGQSTAHPP-----APAHH 479
QY 1312 DMWEGRVGRATSSASIEGLMGRAPPERHSPHLKEQHIRGSIQTGPISYV---VBAQE 1368
Db 480 HQQQ-----QQPQOPQOQH-HGNSGPPPPGAYPIPLESSN 516
QY 1369 DYLRREAKLKREGTPPPPPSRDLTEAYKTALGLPKLKAHEGLNATVKEAGRSIHEI 1428
Db 517 SHHAHPYNNPSLGLSRFPYPP-----GPAHLPPSH-QGV-SYSQAG----- 555
QY 1429 PREELRHTPELAPRPLKEGSIQTGTLKVDYTGASTTGSKKHVRSILIGSPGRT--PPP 1486
Db 556 -----PNGP-PVSSSSNSGSSQAAYSCSPSSQ-----GPGASYPPPP 596
QY 1487 VHLDMADARALACRYEESLKSPPGTASSGSGSIARGAPVIVPELGPQSPPLYE-- 1544
Db 597 VPPITTS-----ATLSTVIATVASPGYKTAASPPGPQYSKRAPSPGYSYKTA 645
QY 1545 -----DHGAP-----FAGHLPRGSPVTMREPTPLQEGSLSSSKASQDRKLSTP 1589
Db 646 TPGYKPGSPSFRGTGTPPYRGTSPPAGPGTFKPGSPVPGPLPPAGPSSLSLPPPP 705
QY 1590 REIAKPSHSTVPEHHPHISPYEHLRGVSGVDLYRSHIPLAFDPTISPRGIPLDAAAY 1649
Db 706 RAPTTGPPLTATQIKQEPAAEYE-----TPESPVPAPSPSPPPKVVVD----- 748
QY 1650 YLPHLAPNPTYPHLYPPVILRGYPDTAALENQTIINDYITSQQMHNTATAMAQRADM 1709
Db 749 -VPSHASQAR-----FNKHLDRGFNSCA-----RSDL 775
QY 1710 LRGLSPRESSLALNAAAGPRGIIDLSQVPHLPVLVPTGTATAMDRLAYLPTAPQPS 1769
Db 776 Y--FVPLEGS-----KLAKGRADLVEKVRREAE 801
QY 1770 SRHSSSPLSPGPGTHLTKPTTTSSERERDRDRDREREKSIILTSTTIVE--HAPIW 1827
Db 802 QR-----AREEKEREKEREKEREKEREKERSVKLAQEGRAPVE 843

QY 1828 RPTEQSSGSSGSSGGGGSSSRPASHAHQHSPISPRTQDALQORPSVHLNTHMGKII 1887
Db 844 CP-----SLGEPVPH-----RPPPEPGS--AVATVPPYL-GPDTPALR 877
QY 1888 TAVEPSKPTVLRSTSTSSPVRPAATFPATHCHPLG-----GTLDGVVPTLMPEVLLPKEAP 1943
Db 878 TLSEYARPHVMSPGNRNHPF-----YVPLGAVDPGLLGYNVNPPALYSSDPAARE 927
QY 1944 RVARPERPADTCHAFILAKPPARSGLPASPSSKSGSEPRPLVPPVSGHATIAARTPAKNL 2002
Db 928 REAREDRDLRLKPGFEVXP---SELEPLHGVPGGLDP---FPHGGLALQPGPPGLHP 981
QY 2003 APHASDPDP-----PAPPASADPHREKTSKPFQIQLERLSLGYHSSYSPEGVE 2054
Db 982 FPHPSLPLGERERLALAAAGPALRPD-----MSYAE-RLAERQHAERVAALGND 1030
QY 2055 PVSFVSPSLTHDKGLPKHLEELDLSHLEGELEPKQPGPVKLGEAAHLPHLRPLPESQP 2114
Db 1031 FLARQLQMLNT-----PHHHQH---SHIHSHLHQODAIHAASASVH-PLIDPL----- 1076
QY 2115 SSSPLQLTAPGVKGHORVVTLAQ-----HISEVITQDYTRHH-----PQOLSAP 2158
Db 1077 ASSHLTRIPYPAG-----TLNPLPLPHPLHENEVL-----RHQFAAPYRDLPASLSAP 1126
QY 2159 LPAPLYSPFGASCPEVLDLRRPPSD---LYLPPPDHGA--PARGSPHSEGGKRSPEP 2209
Db 1127 MSA-AHQLOAMHAQSAELQRLALEQQOWLHAHPLHSVPLPAQEDYVYSHLKESDKP 1182
RESULT 42
ABF1 MOUSE
ID ABF1 MOUSE STANDARD; PRT; 3726 AA.
AC O61329;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)
DE (AT-binding transcription factor 1).
GN ATBF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN NCBI [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/MK X ICR; TISSUE=Brain;
RX MEDLINE=96194902; PubMed=8654949;
RA Igo A., Miura Y., Watanabe M., Sakai M., Inoue Y., Miki T.,
RA Hashimoto T., Morinaga T., Nishi S., Tamaoki T.,
RT "Cloning of the cDNA encoding the mouse ATBF1 transcription factor.";
RL Gene 168:227-231(1996).
RN [2]
RP INTERACTION WITH FNBp3
RX MEDLINE=97315177; PubMed=9171351;
RA Bedford M.T., Chan D.C., Leder P.;
RT "FAP WW domains and the Abl SH3 domain bind to a specific class of
RT proline-rich ligands.";
RL EMBO J. 16:2376-2383(1997).
CC -!- FUNCTION: Transcriptional activator that binds to the AT-rich core
CC sequence of the enhancer element of the AFP gene.
CC -!- SUBUNIT: Interacts with FNBp3.
CC -!- SIMILARITY: Contains 4 homeobox domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D26046; BAA05046.1; -;
CC

Db	2480	-KUPQP-----AAPSLLPQPPQAPPPQCPQLPQSPSPSQLSHLPLKPLHTSTP-----	2522
Qy	1037	AEAQKLPGDPP-----CMTSGLPFPVPPPREVIKAGHPADPSAFSAPPD-----HP----	1083
Db	2527	---QQLANLFPQLIPYCDCKLAF-----SFEHQEHOQLHFLSNQOFIHPQFL	2575
Qy	1084	-----LPLGLHDTARVPLRPPTTISNPPPLI---SSAKHPS-----VLEROI---GAIS	1126
Db	2576	DRSLDMFMLFDPNPLL-ASQLLSGAIPOIPASSATSPSTPTSTWNTLKRKLEEKASAS	2634
Qy	1127	QGMS-----VOLHVPYSEHAKAPVGPVTMGGLPMDPKK-----	1160
Db	2635	PGENDSGTGGEFPQDKRLRTTTITPEOLEILYQYL-----LDSNTRKMLDH	2682
Qy	1161	LAPSGVKQBOL-----SPRGQ---AGPPES-----LGVPATAOE-----	1192
Db	2683	IAHEVGLKRVQVWFONTARERKGFRAVQAHRRCPPCRALFKAKTALEAHISR	2742
Qy	1193	-----SVLRGTALGSV-PGGSITKGIP-----	1213
Db	2743	HWEAKEAGYNLTLSAMLLDCDGLQMKGDIFDGTFSFSLHPSSSDGQGVLPSPVSKTME	2802
Qy	1214	-----STRVPSDSAITYRGSTHGTTPADVLYK	1240
Db	2803	LSPRTLSPSSIKEVGIEDFESPMSGVNLNFDQTKLDNDCCSVNTAITDTTTGD-----E	2859
Qy	1241	G-----TITRIIGEDSPSLDRG-----REDSLPGKHV-----IYE---GKKGH	1276
Db	2860	GNADNSATGIATETKTSAPNEGLTKAAMAMSEYEDRSSGLVSPAPSYKEYDNEGT	2919
Qy	1277	VLSEYEGMSVTQCKEDGRS-----SSGPPHETAAPK-----RTV	1311
Db	2920	VDYSETSSLADPCSPSGSAGSKSGDGDGRPQKRFRTQMTNLQVLKSCFENDYRTP	2979
Qy	1312	DMMEGRV-----GRA-----TSSASIEG-----LMGR	1333
Db	2980	TMLECEVLGNDIGLPKRVQVWFQNAKAKEKSKLSMAKHFGINQTSYEGPKTECTLGI	3039
Qy	1334	AIPPERHSPHLKEOHIRSIQTGIPRSVVEAQEDYLREAKLLKREGTPPPPPSRDL	1393
Db	3040	KYSARLSVRDHIIFSQQHI--SKVKDTIGSQLDKEKEYF-----DPATVRQL	3083
Qy	1394	TEAYKTQALGFLKLPKPAHEGLVATVKEAGRSIHEIPREELR---HTPELPLAPRPLKEGS	1450
Db	3084	MAQELD-----RIKANEVLGLAAOQOG-MFDNAPLOALNLTPVTPALOGIPVLLPGL	3137
Qy	1451	ITOGTPLYDTGASTGSKKHVRSLLIGSPGRTPP-PVHPLDVMADARALERACYEESLK	1509
Db	3138	NRPSLPGFTPANTALTSPKE---NLMLGPSTTVPSGPLP-----TSGLP	3178
Qy	1510	SRPGTASSSGSITARGAPVIVPELGRKPROSPLYEDHGAPFAGHLPRGSPVTWRETPRL	1569
Db	3179	NKPSASLSPTFAQATMAWAPPPPOPPQPPVQPPPP-----PYAAQOIPAPOTPOQ	3234
Qy	1570	QEGSLSSSKASQDKL---TSTPREIAKSPHSTVPHEHPHPIS-PYEHLLLRGVSGVDLYR	1625
Db	3235	QRKDKGEGKEKEKAHKGGEPLVPVKKEKGAPAGTGTISAPLPAMEYAVDPAQLQA	3294
Qy	1626	SHIPLAFDPTS-----IPRGIPLDAAAYVLPRLHAPNPTYP-HLYPPPYLIRG-YPDTA	1677
Db	3295	LQAALTSDPTALTTSQFLPVFPV--GFSYYAFO--IPGALQSGYLQPMYMGMEGLFPYSP	3350
Qy	1678	AL-----ENRQIINDIYIITSQWHTTATAMAQRADMLRGLSPRESSLA	1721
Db	3351	ALSRPLMGLSPGSLLOQYQOYQSLOSBAIQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ	3410
Qy	1722	LNVAAGPRGIIDLQSVPHLPVLPVPTTGTATAMDRLAYLTAPOPFSSRH---SSSPLS	1778
Db	3411	VQOQOQOQOQPKASQTP-----VPOGAASFDKPAK-----ESKPEEQKNVPRELSPLL	3460
Qy	1779	PGGPTHLTKPTTSSSRRERD-----RDREDRDRERE-----KS	1813

```

Db 3461 PKPPEPEABSKASADSLCDPFIIVKQYKLVCRKQAGFGDEEABRSHLKSCLCFGGQS 352d
Qy 1814 ILTSTTTVEHAPITWRPGTEQSSGSSGSSGGGGSSRRPASHSHAHQHSFISPRTDALQQ 1873
Db 3521 VVNLQEWLVHP---TSGSGGGGGGGGGGGGSGYHCLACESALCGEEALSQHEALSALHK 3577
Qy 1874 RPSVLNHTGKGIITAVEPSKPTVLKSTSTSSVPRPAATFPPTHCPGLGTLGTVPTLM 1933
Db 3578 -----HRTITRAARNAKE----- 3590
Qy 1934 EPVLLPKPEAPRVARPERPRADT---GHAFKAPPARSGLEPASPSKSGSEP---RPLVP 1986
Db 3591 HPSELLPHSA---CFPPDPSTASTSQSAAHNSNDSPPP-----PSAAPSSSASPASHKSWP 3641
Qy 1987 PVSCHATIARTPAKNLAPHASDPDPAPASADPHREKTQSPFSTQIELELSRSLGYHGS 2046
Db 3642 PVGSRASAAPPS-----FPPLSSSST-----VTSS 3667
Qy 2047 SYSPGEGVEPVSPVSSPSLTHDKGLPKHLEBDKSHLEGEIRPKQGPVKLGG 2098
Db 3668 SCSTSGVQPSMPTDDYSESDTDLQOK-SDGPASPVPEG---PKDPSPCKDSG 3715

RESULT 43
CA24 ASCSU
ID CA24 ASCSU STANDARD; PRT; 1763 AA.
AC P27393;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
OS Ascaris suum (pig roundworm) (Ascaris lumbricoides).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC Ascariidae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS I AND II).
RX MEDLINE=91340768; PubMed=1714907;
RA Pettitt J., Kingston I.B.;
RT "The complete primary structure of a nematode alpha 2(IV) collagen
  and the partial structural organization of its gene.";
RL J. Biol. Chem. 266:16149-16156(1991).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=I;
CC IsoId=P27393-1; Sequence=Displayed;
CC Name=II;
CC IsoId=P27393-2; Sequence=VSP_001159;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```


Db 1461 -----GLPGLDGLPPSG-----PPGAGAKGRDGPFGPGMPGKGPAP----- 1499

QY 2213 SVLGG--GEDGIEPVSPGEGTEPHGHSASVYPLLYRDG-----EOTESRNGSKSPGNTS 2266

Db 1500 ----GLPFGPGIEGIPGPGIPGPGSPGPGP--SYKDGFLLVKHSOTSEVPQCPGPMVK 1554

QY 2267 QPPAFESKLTESNASVSKKQ-----EINKLNTNRNEPEYNI 2306

Db 1555 LWDGSLYLIEGNE---KSHQDLGHAGSCLSRSTMPFLFCDDVNNVCNYSARNDKSYWL 1611

QY 2307 SQPGTEIFNPAITGTGLMYTSQAVQEHASTNMGLEAIRKALMGKYDQWEESPLPSAN 2366

Db 1612 ST--TAPIPMVPSGGIEPIVSRCAVCEAPAN-----VIAVHSQTIQIPN-CPN 1658

QY 2367 AFNPLNASASIPAAPIYTAADORSNHTLTSFGGGKAKVSRPSRKAKSPAPGLASGDR 2426

Db 1659 GWNLSIWIGYSF--AM-----HTGAGAEQGGQS-----LSSPGSCLEDPA 1696

QY 2427 PPSVSSVHSEGDEN 2440

Db 1697 TPFIECNGARGTCH 1710

RESULT 44

DRPL_HUMAN

ID DRPL_HUMAN STANDARD; PRT; 1185 AA.

AC P54259; Q99495; Q99621; Q9UEK7;

DT 01-OCT-1996 (Rel. 34, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Atrophin-1 (dentatorubral-pallidoluysian atrophy protein).

GN DRPLA.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain, and Cerebellum;

RX MEDLINE=95144175; PubMed=7842016;

RA Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K., Inoue T., Yamada M.;

RA "Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA)."; Nat. Genet. 8:177-182(1994).

RL [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=96026098; PubMed=7485154;

RA Onodera O., Oyake M., Takano H., Ikeuchi T., Igarashi S., Taji S.;

RA "Molecular cloning of a full-length cDNA for dentatorubral-pallidoluysian atrophy and regional expressions of the expand alleles in the CNS."; Am. J. Hum. Genet. 57:1050-1060(1995).

RL [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=96262314; PubMed=8965642;

RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C., Kidwai A.S., Ashworth R.G., Ross C.A.;

RA "DRPLA gene (atrophin-1) sequence and mRNA expression in human brain."; Brain Res. Mol. Brain Res. 36:219-226(1996).

RL [4]

RP SEQUENCE FROM N.A., AND POLYMORPHISM OF POLY-GLN REGION.

RC TISSUE=Brain;

RX MEDLINE=97228904; PubMed=9074930;

RA Ansari-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;

RA "Large-scale sequencing in human chromosome 12p13: experimental and computational gene structure determination."; Genome Res. 7:268-280(1997).

RL [5]

RN [5]

RP SEQUENCE OF 1-76 FROM N.A.

RX MEDLINE=97005364; PubMed=8852663;

Yanagisawa H., Fujii K., Nagafuchi S., Nakahori Y., Nakagome Y., Akane A., Nakamura M., Sano A., Komure O., Kondo I., Jin D.K., Soerensen S.A., Porter N.T., Young S.R., Nakamura K., Nukina N., Nagao Y., Tadokoro K., Okuyama T., Miyashita T., Inoue T., Kanazawa I., Yamada M.;

"A unique origin and multistep process for the generation of expanded DRPLA triplet repeats"; Hum. Mol. Genet. 5:373-379(1996).

[6]

SEQUENCE OF 470-725 FROM N.A.

TISSUE=Brain cortex;

MEDLINE=93315145; PubMed=8325628;

Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;

"Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms."; Genomics 16:572-579(1993).

[7]

INTERACTION WITH WWP1 AND WWP2.

MEDLINE=98313405; PubMed=9647693;

Wood J.D., Yuan J., Margolis R.L., Colomer V., Duan K., Kushi J., Kaminsky Z., Kleiderlein J.J. Jr., Sharp A.H., Ross C.A.;

"Atrophin-1, the DRPLA gene product, interacts with two families of WW domain-containing proteins."; Mol. Cell. Neurosci. 11:149-160(1998).

-!- SUBUNIT: Interacts with WWP1 and WWP2.

-!- TISSUE SPECIFICITY: Relatively high levels in the brain, ovary, testis and prostate. Lower levels in the liver, thymus and leukocytes.

-!- POLYMORPHISM: The poly-Gln region of DRPLA is highly polymorphic (7 to 23 repeats) in the normal population and is expanded to about 49-75 repeats in DRPLA patients. Longer expansions result in earlier onset and more severe clinical manifestations of the disease.

-!- DISEASE: Defects in DRPLA are the cause of dentatorubral-pallidoluysian atrophy (DRPLA) [MIM:125370], an autosomal dominant neurodegenerative disorder characterized by a loss of neurons in the dentate nucleus, rubrum, globus pallidus and Luys' body. Clinical features are myoclonus epilepsy, dementia, and cerebellar ataxia. Onset of the disease occurs usually in the second decade of life and death in the fourth.

-!- CAUTION: Ref.2 sequence differs from that shown due to several frameshifts.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D31840; BAA06626.1; -

EMBL; D38529; BAA07534.1; ALT_FRAME.

EMBL; U23851; AAB50276.1; -

EMBL; U47924; AAB51321.1; -

EMBL; D63808; BAA23631.1; -

EMBL; L10377; -; NOT_ANNOTATED_CDS.

PIR; G01763; G01763.

Genew; HGNC:3033; DRPLA.

MIM; 607462; -

MIM; 125370; -

GO; GO:0005717; Cytoplasm; TAS.

GO; GO:0005634; C:nucleus; TAS.

GO; GO:0005515; F:protein binding; IPI.

GO; GO:0007417; P:central nervous system development; TAS.

InterPro; IPR002951; Atrophin.

Pfam; PF03154; Atrophin-1; 2.

PRINTS; PR01222; ATROPHIN.

Triplet repeat expansion; Polymorphism; Epilepsy.

DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE).

DOMAIN 302 305 POLY-PRO.

DOMAIN 376 382 POLY-SER.

DOMAIN 386 397 POLY-SER.

FT DOMAIN 442 447 POLY-PRO.
FT DOMAIN 483 POLY-HIS.
FT DOMAIN 484 POLY-GLN.
FT DOMAIN 504 POLY-PRO.
FT DOMAIN 564 POLY-SER.
FT DOMAIN 704 POLY-PRO.
FT DOMAIN 802 ARG/ALA-RICH (MIXED CHARGE).
FT DOMAIN 816 ARG/GLU-RICH (MIXED CHARGE).
FT DOMAIN 925 ARG/GLU-RICH (MIXED CHARGE).
FT CONFLICT 94 MISSING (IN REF. 3).
FT CONFLICT 333 H -> Y (IN REF. 1).
FT CONFLICT 339 M -> I (IN REF. 3).
FT CONFLICT 541 P -> T (IN REF. 6).
FT CONFLICT 1028 M -> G (IN REF. 1).
SQ SEQUENCE 1185 AA; 124773 MW; 5135744CEE491C18 CRC64;

Query Match 2.8%; Score 368; DB 1; Length 1185;
Best Local Similarity 21.0%; Pred. No. 2.9e-05;
Matches 327; Conservative 139; Mismatches 604; Indels 490; Gaps 74;

QY 762 TQNGKPPATLIGADGPPGPPPTPRRTSRAPIEPTPAEATGAPTPPPAPSPSAPPV 821
DB 3 TRONKDSMSGRKKEAPGREELRSGRASPGCVSTSSDG----- 45

QY 822 VPKEKEEETAAAPPVREG---EEOKPPAAELAVDTGKAEPEVKSECTEAEAGPAKGK 878
DB 46 --KAEKSRQAKARVEASTPKVNKQGRSEISESEETNAPKTKTEQELPRPQSPS 103

QY 879 DAAEAETAGALKAEKKEGSGRATTAKSGAPQSDSSATCSADEV-----DE 928
DB 104 DLDSLD-----GRSLNDGSDSDPRDIDQDNRESTSPISYFGSVEVNDSDS 147

QY 929 AEGGDKNLLSPRLSLPTGDPANASPOKPLDLKQKRAAALPPIQVTKVHEPPRED 988
DB 148 SSGLSQGPARYHPPLPPFPSPQDPDSTRQP-----EASEFHPSTVTCYHAPMEPP 201

QY 989 AAPTKEAPP-APPPQNLOPE-----SDAPQOQG-----SSPKRSRSPAPPADKE 1033
DB 202 TSMFOAPPAGAPPHPPLQYPGTGVLGGPMPGKGAASSVGGPFGKQHPPTTPTIS 261

QY 1034 AFABEAQKLGDPDCWTSGLPFPVPPREVIKAS--PHAPDPSASYPAPGHLPLGLHDT 1091
DB 262 VSSGA-----SGAPPTKPTPTTPVGGNLPSPAPPANFPHVTPNLP----- 302

QY 1092 ARPVLPRPTTISNPPPLSSAKHPSVLEROIGALISQMSVOLHVPYSEHAKAPCVPTMG 1151
DB 303 -----PPALRLNNAASAPGLCA-----QPLGHLPL-SHAMQO-----GNG 340

QY 1152 LPLMPDKK---LAPFGVKQQLSPRQAGPPESLGYPTAQEAASVLRGTALGSVPGGSI 1208
DB 341 -GLPPGPEKGTLPAP---SPHSLPPASSAPAPMPRFPYSSSSS--SSAAASSSSSSSS 393

QY 1209 TKGIPSTRVPSDAITYRGSITHTGTPADVLYKGTITRIIGEDSPSRLDRGREDSLPKGHV 1268
DB 394 SSASP---FPASQALP---SYPHSPPPP-----TSLSVSNQPPKY---TQPSLPSQAV 437

QY 1269 IYEGKGHVLSEGGMSVTQCKEDGRSSGPPHETAPKRTYDMWEGRVGRAISSASIE 1328
DB 438 -----WSQGGP-----PPPPY-----GRLLNNAH 458

QY 1329 -----GLMGAIPPRHSPHLKQHHIRGSIQTGIPRSYVEAQEDYLRRRAKLKR 1380
DB 459 PGPPPPSTGAOSTAHPV--STHH-----HH-----HQQQQQQQQQQQQH 498

QY 1381 EGTTPPPPPSRDLTEAYKTQALPKLKAHEGLVATVKEAGRSIHETPRE-----ELR 1434
DB 499 HGNSGPPPP-----GAPPHLEGSSHHAHPYAMSPSLGSLR 535

QY 1435 HTPLEPL-APRLKEGSITQ---GTPLKYDTGASTTGSK-----KHDVRSI--IGSPGR 1482
DB 536 PYPPGPAHLPPHPHQVSQAGPPPVSSSSNSSSTSSQSYPCSHFSPSQGQAP-Y 594

QY 1483 TFPVPHPLVDNADARALERACVEESLKRPGTASSSGGSIARGAPVIVPELGKPRQSPLT 1542

Db 595 PFPVP-PTVTSSA-----TLSTVIATVASSPAGYKTASPPGPPYCKRAPSPCA 643
QY 1543 YEDGAPFAGHLPRGSPVTMREPTPRLQEGSLSSSKASQDRKLTSTPREIAK-SPHSTVP 1601
Db 644 YKTATPP--GYKP-GSPSPFRGTGPPGYRG-----TSPGAPGTGFKGS--P 685
QY 1602 EHHPHIPSPYBHLIRGVSGVDLYRSHIPLAIFDPTSPRGIPDLDAAYVYLPRHLAPNTY 1661
Db 686 TVGPGPL-----PPAGPSGLP-----SLPPPPAA 709
QY 1662 PHLYPPVILIRGYPDTAALENQTIINDYITSQMHNTATATAQADMLRGLSPRESSLA 1721
Db 710 PASGPP-----LSATQIKQEPABEYET----- 731
QY 1722 LNYAAGRGIIDLQVPHLPVLPPTGCTPATAMDRLAYLPTAPOPFSS--RHSSSPLSPG 1780
Db 732 -----PESVPPARSFPP--PKVDVPSHSAQSARFNKHLDRGFNSCARSDLYFVPL 782
QY 1781 GPTHLT-----PTTSSSERDRDRDREREKESILTSTTVE--HA 1824
Db 783 EGSKLAKRADLVKVRREAQRAEERERERERERERERERERERERERERERERERER 842
QY 1825 PIWRPGTEQSSGSGSGSGSGSGSRPASHAHQHSPISPTQDALQORPSVHLNTGMK 1884
Db 843 PVECP-----SLGVPVPH-----RPPFEPGS--AVATVPPYL-GPDTTP 876
QY 1885 GIITAVBESKPTVLRSTSTSSVPRPAATFFPATHCPLG-----GTLDGVPVPTLMFVLLPK 1940
Db 877 ALRITSEYARPHVMSVGNRHPF-----YVPLGAVDFGLLGYNNVPALYSSDDPAAR 926
QY 1941 EAPRVAPERPRADTGHAFKAPPARSGLEPASS--PSKSGEPRPLVPVPSGHATTIARTPA 1999
Db 927 ERESEARERDLRLKPCFEVKP---SELEPLHGVPGGGLDP---FPRHGGALQPGPG 980
QY 2000 KNLAHPHASPP-----PAPPASADPHREKTQSKPFSIQEILELSLGVHSGSYSP 2051
Db 981 LHPFPFHPSLGLERERLALAAAGPALRPD-----MSYAE--RLAAERQHAERVAAL 1029
QY 2052 GVEVPSVSSPSLTHDKLPHLELDKSHLEGELPKQPGPVKLGGEAAHLPHLRPLPE 2111
Db 1030 GNDPLARLQMLNVT-----PHHGH--SHTHSHLHLHQDAIHAASASVH-FLIDPL-- 1078
QY 2112 SQSSSPLLQAPGVKQHVVTIAQ-----HISEVITQDYTRHH-----POOL 2155
Db 1079 ---ASGSHLTPYPAG-----TLNPLPHPLHENEVL-----RHQLFAAPYRDLPASL 1125
QY 2156 SAPLAPLYSPGASCPVLDLRRPSPD---LYLPPDPHGA--PARGSPHSEGGKRSPEP 2209
Db 1126 SAPMSA-AHQAMHAQSAELQRLALEQQQWLHAHPLHLSVPLPAQEDYYSHLKKESDKP 1184

RESULT 45
BCL9 HUMAN STANDARD; PRT; 1426 AA.
ID BCL9 HUMAN AC 000512;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).
GN BCL9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98158621; PubMed=9490669;
RA Willis T.G., Zaiberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,
RA Jaydel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,
RA Dyer M.J.S.;
RT "Molecular cloning of translocation t(1;14) (q21;q32) defines a novel

Db 1009 MMSRMSKAMPSSSTLYHDAIKTAVASDDDDSPARS--PNLPSMNMNP-----GNG 1057
QY 2043 YHGSSYEGVEPVSVSPSLTHDKLPKHLLELDKSHLEGELPKPGVPLGGEAAH 2102
Db 1058 INTQNPRISSGNPVVPM--PTLS-PWGMTQPL-----SHSN-----QMPSPNAV----- 1099
QY 2103 LPHRLPLPESQSSSPLLOTPAGVKGHQVTVTLAQHISEVITQDTRHHPOOLSAPLPAP 2162
Db 1100 -----PNIPPHGVPM--GFGLMHNPIIM-----GHGSGEPMPVPG 1133
QY 2163 LYSPPGASCPVLDRPPSDLYLPPDHGAPARGSPHSEGG-----KESPEPN 2210
Db 1134 RMGFPQGGPPV--QSPQVQVFP-----HNGFSGGGSPGPGFPGEGPLGRPSNLPSS 1187
QY 2211 KTSVL--GGGSDGIEPVPPGEMTBFHSHSAVYPLLYRDGEQTBPSSRMGSKSPGNTSQP 2268
Db 1188 ADAALCKPGGPG-----PDSFTVLGNSMPSVF-----TDPDLQEVIRPGATGIP 1232
QY 2269 PAFPSKLTESNSAMVSKQBEINKKLTNNRNEPYNISQGTETIFNNPAITGTGLMTYR 2328
Db 1233 EFDLSRIIPSEK--PSQTLQVFRGEVGRKQPO-----GPGPGFSHMGMG----- 1278
QY 2329 SQAVQEHASTWGLEAIRKALMGKYDQWESPPLSANAFNPLNASLSLPAAMPIT----- 2384
Db 1279 -----EOAPRMGL-----ALPGM-----GGPGVGTPTDPLGTAPSMPCGHNMRPPAF 1321
QY 2385 AADGR--SDHTLTSPGGGKAKVGRSSRSRAKSPAGLAGS-DRPPSVSSVHSEGCNR 2441
Db 1322 LQCGMGMGHHRMSP--AQSTMPCOFTLMSNPAAVGMIFGKDRGAGLTYH----- 1371
QY 2442 RTPLTNRWEDRPSAGSTPFYFNPLMLRQAGVNAS-----PP--PPGLPAGSG 2489
Db 1372 -----PGPVGS-----PGMWSMQGMGMPQONIMIPPQMRPGMAADVG 1410

RESULT 46
GSRI HUMAN STANDARD; PRT; 1509 AA.
AC Q9NZM4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glioma tumor suppressor candidate region gene 1 protein.
GN GLTSCR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
RA Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
RA Scheithauer B.W., Louis D.N., Jenkins R.B.;
RT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
region."
RL Genomics 64:44-50(2000).
CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
CC placenta, skeletal muscle, and pancreas, and at lower levels in
CC lung, liver, and kidney.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.iesb-sib.ch/](http://www.iesb-sib.ch/announce/)
CC or send an email to license@iesb-sib.ch).
CC -----
CC EMBL; AF182077; AAF62874.1; --
DR Genew; HGNC:4332; GLTSCR1.
DR MM; 605690; --
FT DOMAIN 37 POLY-GLY.

FT DOMAIN 884 889 POLY-PRO.
FT DOMAIN 1214 1225 POLY-SER.
FT DOMAIN 1282 1286 POLY-PRO.
FT DOMAIN 1294 1304 POLY-PRO.
SQ SEQUENCE 1509 AA; 152991 MW; 7C5144F443CE6821 CRC64;
Query Match 2.7%; Score 360; DB 1; Length 1509;
Best Local Similarity 23.0%; Pred. No. 6.4e-05;
Matches 249; Conservative 104; Mismatches 420; Indels 308; Gaps 51;
QY 745 SSDTESIPSPHTEAAKDTGQNGPKPPATLGADGPPPG---PPTPRRTSRA-----PIPT 797
Db 514 SLPTQSQAPAPAGPAATTVLQGVTLPPSAVAMLNTDGLVQVPAATPAATGAAAPVITVQPA 573
QY 798 P-ASEATGATP-----PPAPPSP-SAPPPVVPKEEKEEETAAAPVVEEGEOKP 845
Db 574 PQAPPAVSTPLPLGLQOQPAOQPPQAATPQAAPQAATPQPSGLASSP--EKIVLQGP 631
QY 846 PAABELAVDTG-----KAEFVKSECTEABEGPAKGDAAEABATAG 889
Db 632 PSATPTAILTQDSLQMFLPQERSQOPLSAEGPHLSVPASVIVSAPPAQDPAPATPAVAKG 691
QY 890 ALKAEKEGSGRATTAKSSCAPQDSSA-----TCSADEVDEAEGDKNLLSPR 941
Db 692 A-----GLGQAPDQASAPAPAPQIPAAAPLKGPGSSPSLPHQAPLGDSPHLPSPH 744
QY 942 PSL--LPTGTGPRANASP--OKPLDLKQLKORAAAIPIQVTKVH--EPPREDAATPKP 994
Db 745 PTRPSPRPSPQSVSRPPSEPLHPCPPQAPPTLPGIFVIONQLGVPPPASNPAPTAP 804
QY 995 APPAPP--PPQNLQESDAPQOP-----GSSPRGKSRSPAP-PADKEAFAAEAAQ- 1040
Db 805 GPPQPLRPQSPPEGPLPPAPPLPPSPSTSSAVASSSETSSRLPAPTSDFLQPPSQG 864
QY 1041 --KLPGDPPCW-----TSGLPFPVPPRE--VIKASPHAPDPSA-----FSYAPPGHPLPLG 1087
Db 865 PHKSFTPTTLHLVPEPAAPPPPPPTFMVTTTFFPALPQKALLERHQVPSG---II 920
QY 1088 LHDTRPVLPVLPPTISNPPPLISSAKHPSVLEROIGALSQMSVQLHVPYSEHAKAPVP 1147
Db 921 LQKAGAPAAPQTSLSGLTSPA--ASVL-----VSGQAPSGTPTAP--SHAAPAPM 971
QY 1148 VTMLGP--LPMDPKLAFFSGVQEQLSPRQAGPPESLGVPYTAQESVLRTALGSVPG 1205
Db 972 AATGLPPLPAENKAFA--SNLPTLNKAKAASSGPGKPSGLQYESKLSGLK- 1021
QY 1206 GSITKGPSTRTVPDSAL-----TYRGSITH-----GTPA--DVL-----YKGTI- 1243
Db 1022 -----PPTLQPSKEACFLEHLKHOGSVLHPDYKTAFPSFEDALHRLPHYVYQAGALP 1074
QY 1244 -----TRIIGED-----SPS----- 1253
Db 1075 SPDSYKHVDEBEFETVSTQLKRTQAMLNKYLLLEBSRRVSPSAEMWIDRMFTQSEKT 1134
QY 1254 --RLDRGREDSLPKGHV-----IYEGKKGHVLSYEGGMSVTQCSKEBGRSSSGPPH 1302
Db 1135 TLALDKLQAKKEPKDEYVSSSSSLGLPTAASSEGHRLPGHGLS---SSAPGASTQPPPH 1190
QY 1303 -----ETAAPKRTYDMMEGRVGRALISSASIEGLMGRAPTPPERKSHPLKQHHI 1351
Db 1191 LPTKLVIHGGAGGSPSVTW-----ARASSLSLSSSSSSSSAASLSDADGDGMPSPNRPP 1246
QY 1352 RGSITQIPRSYVAQEDY-----LRREAKL-----LKREGTPPPPPSRDLTEAYK 1398
Db 1247 -----KTY--BARSIGLKLKQKAGLSKVVNHTALDPVHQPPPPATLKVKEP- 1294
QY 1399 TQALGPLKLKPAHEGLVATVKEAGRSIHEIPREBLR-----HTPELPL----- 1441
Db 1295 -----PRPPPPPP-PTGQMGCTVDHPHPAAPEKPELGTATPHCPRLPLAKTYRENVGG 1346
QY 1442 --APRPLKBSITQGT--PLKYDTGASTTSGKHHDVRSLLIGSPGRTTFFPHPLDVMADAR 1497
Db 1347 PGAPEGTAGARGGSPAPLPKAVDEATSG-----LIRELA 1382

QY 1498 ALERACYEESLKRSP-----GTASSGGSTARGAPVIVPELCKPROSPLTYEDHGAPFAGH 1553
 Db 1393 AVEDLYQMLKPPPEPAASAAQTQDWDWEAPGLPPAKRKSESPP---DVDQASFSDD 1439
 QY 1554 LPRGSPVMTREPTPLRQBSLSSSSKASQDRKLITSTPREI-AKSPHSTVPEHHHPISPYE 1612
 Db 1440 SPQDD--TLTEHLQSAIDSLNLQAPGRTPAPSYPHAASAGTPASPPPLHRPEAYPSS 1497
 QY 1613 H 1613
 Db 1498 H 1498

RESULT 47
 NEST_HUMAN
 ID NEST_HUMAN STANDARD; PRT; 1618 AA.
 AC P48681; O00552;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nestin.
 GN NES.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=93123384; PubMed=1478958;
 RA Dahlstrand J., McKay R.D.G., Zimmerman L.B., Lendahl U.;
 RT "Characterization of the human nestin gene reveals a close
 evolutionary relationship to neurofilaments."
 RL J. Cell Sci. 103:589-597(1992).
 RN [2]
 RP SEQUENCE OF 296-309 FROM N.A.
 RX MEDLINE=99117355; PubMed=9917366;
 RA Yaworsky P.J., Kappen C.;
 RT "Heterogeneity of neural progenitor cells revealed by enhancers in
 the nestin gene."
 RL Dev. Biol. 205:309-321(1999).
 CC -!- TISSUE SPECIFICITY: CNS STEM CELLS.
 CC -!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
 IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
 CC -!- SIMILARITY: Belongs to the intermediate filament family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; X65964; CRA46780.1; -;
 DR EMBL; AF004335; AAB64426.1; -;
 DR PIR; S21424; S21424.
 DR Genew; HGNC:7756; NES.
 DR MIM; 600915; -;
 DR GO; GO:005882; C:intermediate filament; NAS.
 DR GO; GO:0007417; P:central nervous system development; NAS.
 DR InterPro; IPR001664; IF.
 DR Pfam; PF00038; filament; 1.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neuron.
 FT DOMAIN 1 7 HEAD.
 FT DOMAIN 8 312 ROD.
 FT DOMAIN 313 1618 TAIL.
 FT DOMAIN 8 43 COIL 1A.
 FT DOMAIN 44 55 COIL 1.
 FT DOMAIN 56 151 COIL 1B.
 FT DOMAIN 152 172 LINKER 12.

FT DOMAIN 173 191 COIL 2A.
 FT DOMAIN 192 194 LINKER 2.
 FT DOMAIN 195 312 COIL 2B.
 SQ SEQUENCE 1618 AA; 176704 MW; C9E9AA48C66534D0 CRC64;
 Query Match 2.7%; Score 358; DB 1; Length 1618;
 Best Local Similarity 21.1%; Pred. No. 7.9e-05;
 Matches 323; Conservative 196; Mismatches 511; Indels 502; Gaps 77;
 QY 97 KSEMEFIESKRPL--ELLP--DPLLRPSPL---LATQCPA--GSEDLTKDRSLTGKLEP 147
 Db 329 KLELQFPRTPEGRRLGSLLPVLSPTSLPSPATLETVPDAFLKNQEFLOARTTFLASTP 388
 QY 148 VSPSPPHPTDPELELVPPRLSKEELIQMDRVDREITMVEQOISKL-----KKQOQLE 201
 Db 389 I-PPTQAPSP-----AVDAETIRAQDAPLSLLOTTQGRKQAPFLR 428
 QY 202 EEA--AKP--PEPEKP-----VSP-----PPIESKH-----223
 Db 429 AEARVAIPASVLPGEPEPGQORQEAQSGPDHSLAPLSPDHSSLEAKDGSGGSRV 488
 QY 224 -----RSLVQIYDE---NRKAEAAH-----RILEGL 248
 Db 489 FSICRGEQIMGLVEKETAIGKVSVSLQOEIWEEDLNKKEIQDSQVPLEKETLKS 548
 QY 249 GPQVE---LPLYNQPSDTRYHENIKINQAMRKLLIYFKERNHARKQWKQKFCQRYDQL 305
 Db 549 GEEIQESLKTLENQSHETLE-RENQECPRSLBEDL-----ET 584
 QY 306 MEALEKKVERI-----ENPRRRAKESKVREYEQFPEIRKQERLOERMOSRVQORG 358
 Db 585 LKSLKENKRAIKCGGSEISRKGCQQLKPTGKEDTQTLQ-SLOKENQELMSLEGNLE 643
 QY 359 SGLSMSARSEHEVSEIIDIGLSEQENLEKQMRQAVIPPMYLDADQOQRIKFINNGLMAD 418
 Db 644 TFL-FPGTENQELVSSLQENLESITALEKENQPSRSEV---GDEALRLPTKEN--QE 697
 QY 419 PMKYVDKQVMMWMSQEQEKETER--EKFMQHPKNFGLIASFLERKTVAECVLYVYLTKN 476
 Db 698 PLRSLED-----ENKEAFRLSEKNOEP-----LKTEEDQSIVRPLET 737
 QY 477 ENYKSL-----VRRSYRRGKSGQQQQQQQQQQQQQQQMPRSSQEEKEKEKEAE 530
 Db 738 ENKHSRLSLEEQQDETTLTLEKETQRRRSGLGEQDQWTLRP-----PEKVDLEPLKSLD 791
 QY 531 KEEKEPVENDKEDLLIKEKTDGSDNDEKAVASKRKTANSQGRKRRIETSMANEA 590
 Db 792 QETARP-LENENOEFKLSKEES-----VEAVKSLTEILES-----LKSAGOE 834
 QY 591 NSBEAITPQOSAEIASMELNESSRWTEEM-----ETAKKG-----LLEH 630
 Db 835 NLETLKSPETQAPL-----WTPPEINKSGNESSRKGNSTRITGVCSSPRDIQTP 884
 QY 631 GRNWSAIARVMVGSKTVSQ-----KNFYFNKKQNLDLILQOHKLKMKERNARRKK 683
 Db 885 GRGESGIIETSGSMEPGEFEISRGVDKESQNLNEEENLKGVEQSL-----RSLEE 939
 QY 684 KKAFAAEEAAPPVVEDEME-----ASGVSGNEEENVEEAALHAGNE--VPRGE 735
 Db 940 QELPQSADVQVQWEDTVEKQOELAQESPPGMAGVENKDEAELNREOQGTGKEVEEQE 999
 QY 736 CSGPATVNNSSDTEISIPSP---HTEAAKDTGQNGKPPATL--GADGPPPPGPTPPRTS 790
 Db 1000 LNA-----TEEVWFPGEGHPENPEKQRLVGASVKGAGLQ-----1039
 QY 791 RAPIEPTPASEATGATPPPPAPSPGAPPPVVPKEKEEBETAAPVVEGEEOKPPAAEE 850
 Db 1040 ----DPEGQSQVGTGQLQAPQGLPAIEPLVEDDVAQGDQASPEVMLGSE---PAMGE 1092
 QY 851 LAV-----DTGK-----AEPVSECTE-----EAEEGPAKGDAAEAAT 886
 Db 1093 SAAGAEPGLQGVGLGDPGLHLTREVMPEPPEESLEAKRVQGLEGR---KOLEEA---1147

QY 897 AEGALKARKE-GSGGRATTAKSGAPQDSSATCSADEVDEAR-----GGDKNRLISP 940
Db 1148 -GGIGTEFSELPGKSRDPWPPREGRESEAEAPRGAEAFPAETLHTGSD-----AP 1200
QY 941 RPSLLTPGDPFRANASPKPLDLKOLKORAAIPIQV--TKVHEPPREDAAPTKPAPPA 998
Db 1201 SP-----WPLGSEAEED-----VPPVLVSPGTYTPILLEDA----- 1232
QY 999 PPPQNLOPESDAPQPGSSPRGKSRSPAPPADKEAFAAEAQKLPDGPDPWTSGLPFPVP 1058
Db 1233 -PGLOPQAEQSEASGWQGRABACKVESEBELG-----GEIP-----EGL----- 1274
QY 1059 PREVIKASHPADPSAFYAPGPHPLGLHDTARVLP--PPTISNPPPLISSAKH--- 1114
Db 1275 QEEGESRESEDELTGTLDPSTPLGFLRS---PTSPRWTPLESRGLPKETGKEGWD 1331
QY 1115 PSVL-----ERQIGAISQMSVQLHVPYSE-HAKAPVCPVTMGLPLMDPKKL 1161
Db 1332 PAVLASEGLEPSEKEBEGEGECGRDLSERFEDLGTAPF-----LPGVPGEV 1383
QY 1162 A-PFSGVKQEQLSP-----RQAGPPESLG-- 1185
Db 1384 AEPLGQVQLLLDPAANDRDESDGFADEBSGEGEDQEGREPAGRWGPGSSVGS 1443
QY 1186 -VPTAQASVLRGTALG-SVP-----CGSITKIPSTRVPSDS--AITYRGSITHGTPA 1235
Db 1444 QALSSQGEFLESVSVPWDDSLRGAVA-GAPKTALETESQDSAEPSGSEESDPV 1502
QY 1236 DVLYKGTITRIIGEDSPRLDRGREDSLPKGHVI-----YEGKKGHVLSYEGG-MSV 1286
Db 1503 SLEREDKV-----PGPLEIPSGMEDAGFDGADIIGVNGQGNLKGKSOHV---NGVMNG 1553
QY 1287 TQCSKEDG-----RSSSGPPHETAAPKRTYDMMEGRVGRRAISSAIEGLMGAIP 1336
Db 1554 LEQSEESARNALVSEGRGSPFQEEGSALKR-----SSAG----- 1590
QY 1337 PERHSPHLKBOHIRGSITQIGIPRSYVEAQE 1368
Db 1591 -APVHLGGQQLKFTQREGDRESWSSGD 1618

RESULT 48

APC RAT STANDARD; PRT; 2842 AA.
AC P70478;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein).
GN APC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fischer 344/N; TISSUE=Brain;
RC MEDLINE=96116966; PubMed=8563176;
RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,
RA Sugimura T., Nagao M.;
RT "cDNA cloning of the rat APC gene and assignment to chromosome 18.";
RL Mamm. Genome 6:746-748(1995).
RN [2]
RP MUTAGENESIS.
RC STRAIN=Sprague-Dawley, and Fischer 344/N;
RX MEDLINE=95148647; PubMed=7846077;
RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
RA Weisburger J.H., Sugimura T., Nagao M.;
RT "Specific 5'-GGGA-3'-->5'-GGA-3' mutation of the Apc gene in rat colon
RT tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
CC and participates in Wnt signaling. APC activity is correlated with

CC its phosphorylation state (By similarity).
CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
CC axin (By similarity).
CC -!- PTM: Phosphorylated by GSK3B (By similarity).
CC -!- SIMILARITY: Contains 7 ARM repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D38629; BAA07609.1; -.
CC HSSP; Q02248; 3BCT.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR SMART; SM00185; ARM; 5.
DR PROSITE; PS50176; ARM_REPEAT; 1.
KW Wnt signaling pathway; Anti-oncogene; Phosphorylation; Coiled coil;
KW Repeat.
FT DOMAIN 1 728 LEU-RICH.
FT DOMAIN 1 62 COILED COIL (POTENTIAL).
FT DOMAIN 125 260 COILED COIL (POTENTIAL).
FT REPEAT 451 493 ARM 1.
FT REPEAT 503 545 ARM 2.
FT REPEAT 546 589 ARM 3.
FT REPEAT 590 636 ARM 4.
FT REPEAT 637 681 ARM 5.
FT REPEAT 682 723 ARM 6.
FT REPEAT 724 765 ARM 7.
FT DOMAIN 739 2831 SER-RICH.
FT DOMAIN 1130 1155 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1864 1891 HIGHLY CHARGED.
FT MUTAGEN C->R: IN AN IQ-INDUCED COLON TUMOR.
SQ SEQUENCE 2842 AA; 310530 MW; 3CB2EA8A34E8F47 CRC64;

Query Match 2.7%; Score 355.5; DB 1; Length 2842;

Best Local Similarity 18.7%; Pred. No. 0.00016;

Matches 514; Conservative 334; Mismatches 1019; Indels 879; Gaps 127;

QY 37 VGLLEYQHHRDYASHLSPGSIIOQRRRPRLLSSEFQPGNRSOEL-----HL 84
Db 615 VGLTYRSQTNLAIIEGGGIL---RVNSSLAT---NEDHQILRENNCLQTLLOHL 667
QY 85 R-----PESHSYLPPELGKSEM--EFIESKRPRLELLPDLPLRPS 121
Db 668 KSHSLTIVSNACGTLWNLARNPKQEAALWDMGAVSMLKNLIHSH----- 713
QY 122 PLLATGQPAGEDELTKDRSLTGKLEVPSPSPPHTDPELEL-----VPPRLSKBELIQNM 176
Db 714 KMIAMGSAALRLNLMANR--PAKYKDANIMSPGSLPSLHVVRKQKALEAELDAOHLSETF 771
QY 177 DRVDREITWVEEQIISKLKQOOLEE---BAAKPEPEKPP-----VSP-----P 217
Db 772 DNID-NLSPKASHRSKQKHQNLVGDYVDFASRHHDDNRSDNFTGNMTVLSFYLNTVLP 830
QY 218 PIESKHSVLQIYYDENRKAEEAAHRILEGIGPQVELPLYN----- 258
Db 831 SSSSSRGSU-----DSSRSEKDRSLERERGIG---LSTYHSATENPGTSSKRGQLSAT 881
QY 259 -----OPSDTRYHENIKINQAMRKKLILYFKRRNH--ARKQWK 295
Db 982 AAQIAKYMEEVSALHTSQDQDRSPASAAELHCAVEERAAARRSSASHTHPTNHFAKSSSS 941
QY 296 QKFCQ-----RYDQLEALEKKYERTENNRRRAKESKVREYKEK-----QFP-E 339
Db 942 NRTCSMPYAKVEYKRSSDNLNSTSSDGYGKRGQMKPVSVEYSEDDGKFCSCGYQYAD 1001
QY 340 IRKQRELQRMQSRVQGRGSLMSAARSEHEVSEIIDLSELENLEKQMRQLAVIPML 399

QY 2346 IRKALMGKYDQWEESPLSANAFNPLNASLPAAMPITAAADGRSDHLLTSPGGGKAKV 2405
D 2738 --KAGQSPAPVAETGETCMAERTPFSSSS-----SKH--SSPSGTVAARV 2780
QY 2406 SG--RPSRKA-----PAP-GLASGRPPSVSVHSEG 2437
D 2781 TPFNPNPSRKSSADSTSRPSQIPTVPVGSSTTKRDSKTDSTESSG 2826

RESULT 49
N214_HUMAN
ID N214_HUMAN STANDARD; PRT; 2090 AA.
AC P35658;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nuclear pore complex protein Nup214 (Nucleoporin Nup214) (214 kDa
GN NUP214 OR CAN OR CAIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=92195315; PubMed=1549122;
RA Von Lindern M., Fornero M., Van Baal S., Jaegle M., De Wit T.,
RA Buijs A., Grosveld G.;
RT "The translocation (6;9), associated with a specific subtype of acute
RT myeloid leukemia, results in the fusion of two genes, dek and can,
RT and the expression of a chimeric, leukemia-specific dek-can mRNA";
RL Mol. Cell. Biol. 12:1687-1697(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=94151361; PubMed=8108440;
RA Kraemer D., Wozniak R.W., Blobel G., Radu A.;
RT "The human CAN protein, a putative oncogene product associated with
RT myeloid leukemogenesis, is a nuclear pore complex protein that faces
RT the cytoplasm.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:1519-1523(1994).
CC -!- FUNCTION: May serve as a docking site in the receptor-mediated
CC import of substrates across the nuclear pore complex.
CC -!- SUBUNIT: Homodimer. Interacts with DDx19 and NUP88.
CC -!- SUBCELLULAR LOCATION: Nuclear pore complex. Cytoplasmic filaments.
CC -!- TISSUE SPECIFICITY: Expressed in thymus, spleen, bone marrow, in
CC kidney, brain and testis, but hardly in all other tissues or
CC whole embryos during development.
CC -!- DOMAIN: Contains F-G repeats.
CC -!- PTM: Probably glycosylated as it reacts with wheat germ agglutinin
CC (WGA).
CC -!- DISEASE: Implicated in a subset of acute myeloid leukemia (acute
CC nonlymphocytic leukemia) (AML) carrying a chromosomal
CC translocation t(6;9) (p23;q34) that results in the formation of a
CC DEK-CAN fusion gene.
CC -!- DISEASE: Involved in some cases of acute undifferentiated leukemia
CC (AUL) through a chromosomal translocation t(6;9) (q21;q34.1) that
CC involves NUP214/CAN and SET.
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/CAN.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64228; CAN45535.1; -;
DR PIR; S26058; S26058.
DR Genew; HGNC:8064; NUP214.

DR MIM; 114350; -;
DR GO; GO:0005643; C:nuclear pore; TAS.
DR GO; GO:0005215; P:transporter activity; TAS.
DR InterPro; IPR004325; Nucleoporin_FG.
DR InterPro; IPR001680; WD40.
DR Pfam; PF03093; Nucleoporin_FG; 22.
DR SMART; SM00320; WD40; 2.
KW Nuclear protein; Transport; Proto-oncogene; Chromosomal translocation;
KW Repeat; Glycoprotein.
FT DOMAIN 481 2076 11 X 5 AA APPROXIMATE REPEATS.
FT DOMAIN 1409 2084 18 X 4 AA APPROXIMATE REPEATS.
FT DOMAIN 1427 2085 11 X 3 AA APPROXIMATE REPEATS.
FT DOMAIN 1213 2090 PRO/SER/THR-RICH.
FT DOMAIN 740 768 LEUCINE-ZIPPER 1.
FT DOMAIN 861 882 LEUCINE-ZIPPER 2.
FT SITE 812 813 BREAKPOINT.
SQ SEQUENCE 2090 AA; 213766 MW; 6DBE767FDD857F8F CRC64;
Query Match 2.7%; Score 355; DB 1; Length 2090;
Best Local Similarity 20.6%; Pred. No. 0.00013;
Matches 362; Conservative 177; Mismatches 656; Indels 560; Gaps 79;
QY 993 KPAPPAP-----PPQNLPESD-AFQPGSSPRGSKR 1024
D 381 KTLPPAPVLMLLSTDGVLCPFYMINQNPVGKSLIKTPERLSLEGERQPKSPGTTPTTS 440
QY 1025 SPAP-PADKEAFAAEAKLP-----GPPP 1047
D 441 SQAQKLDASAAAPASLPSPSPAATFSLPAGGATVFSFGSSSLKSSATVTGEP 500
QY 1048 CWTSGLPVPPPREVIKASHAPDPSAFSYAPPGH-----PLPLGLHDHTARVLP 1097
D 501 SYSSG-----SDSSKAAP-GPGSTFSFVPPSKASLAPTP-----AASPVAASAAPS 547
QY 1098 -----RPPTISNPPLIS-----SAKPSVLERQIGA-----ISQMSVQ 1132
D 548 FGSSGFKPTLESTFPVSVSAFNIAKMSFFPSTSAVKYNLSEKFTAAATSTPVSSQSAP 607
QY 1133 LHVPSYSEAK-APVGPVTMGLPLMDPKKLAPFGVKQEQLSPRGQAGPPPSLGVPTAQE 1191
D 608 PMSFPSSASKPAASGPISSHPTPLSAPPSSVPLKSVLP---SPSGRSAGQSSSPVPSMVQ 664
QY 1192 ASVLRTGALSGVPGSGITKGIPTSTRVPSDSAITRGSTHTGTPADVLVYKGTITRIIGDS 1251
D 665 KSP-RITPPAAKPGSPQAKSL-QPAVAEKQGHQWKS-----DPVMAG-----IGEE- 709
QY 1252 PSRLDRGREDLPGKHVIYEGKKHVLSEGMVTCQSKEDG-----RSSSGPPHETAAP 1307
D 710 -----TAHFOKELEELKARTSKACFQVGTSEEMKMLRTESDDDLHTFLLE 753
QY 1308 -KRTYDMMGRVGRGAISSASIEGLMGRAPPERHSPHLKEQHHRGSIITQGI-PRSYVE 1365
D 754 IKETTESLHGDIS-SLKTLLLEGFAG--VEAREQNERNRDSGYLHLLYKPLDPKGS--E 808
QY 1366 AQEDYLREAKLLKREGTPPPPPSRDLTEAYKTQALGPLK-----LKAHEGLVATYKE 1420
D 809 AOLQEIRRLHQYKVF-----AVQDVNDVLDLWDQHLQKQKQKORHLLVPERETLNTLAN 863
QY 1421 AGRSIHIEIPRELRHTTELPLAPPLKE---GSTGTPTLKYDTGASTGKGGKHVRSLLI 1477
D 864 -NREIINQQRKRLNHLVDSLQQLRLYKQTSLWSLSSAVP-----SQSSIHSFDSLSLIC 917
QY 1478 GSPGRTPPPHVPLD--VMADARALERACYEESLKS--PGTASSSGSGSIARGAPVIVPE 1532
D 918 NALKTTIESHTKSLPKVPKLSMPKQQLRNFLAKRTTPVRSPTAPASLSRAEL----- 973
QY 1533 LGKPRQSPLYEDHGAPFAGHLPRGSPVTRMRETPRLQEGSLSSSKASQDKRLTSTPRE- 1591
D 974 -----SQRYEEDL-----DEVSSFSVSQSLESDARTSKCKDEAVVQAPRHA 1016
QY 1592 -IAKSPHSTVPEHHPHISPYP--EHLR-----GVSGVDLYESH---IPLADPTSPISPGI 1641
D 1017 PVVTRP-SIQPSLLPH-AAFPKAKHLVHGSGPGVMGTSVATSKIIIPQADSTML----- 1070

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch).

[illegible]

Qy	1082	HPILPLG-----LHDTARVLPRLRPPTISNPPPLISSAKHPSVLERQIGAISQMSV	1131
Db	415	APGPKNGSGFGAPGNKGDGTGAKGEGPAGVQGP-----GPAGEGKRG-----	459
Qy	1132	QLHVPVSEHAKAPVGPVTWGLPLPMPDKKLAPSGVKVQBQLSPRGQAGPESLGVPTAQE	1191
Db	460	-----ARGPFGA--GLP-----GPAGRGAPGSRGKGFPCAD--	488
Qy	1192	ASVLRGTALGSVPGSITKGIPISTRVPDSAITYRGSITHGTGPADVLYKGTITRIIGES	1251
Db	489	-----GIAGPKGPPCE-----GSPCAVGPKG-----S	511
Qy	1252	PSRLDRGRDRLP--KGHVIEYKKGHVLSYEGCMSVTQCSKEDGR--SSSGPPHETAAPK	1308
Db	512	PGBAGRPGEAGLPCAKGLTSGPSGPGP-----DGKTGPPGADGQPGGPPAGPP-----	560
Qy	1309	RTYDMMEGRVGRAISSASIEGLMG-----RAIPERHSPHHLKEQHHRIGRSITOG	1358
Db	561	-----GARGQA-----GVNGFPQPKGAAGEPKPGERGAP-----G	591
Qy	1359	IPRSYVEAOBYLRLREAKLLKREGTPPPPPSRDLTEAYKTQALGPLKL-----	1407
Db	592	PPCAVGAAGKD--GEA--GAQGPDPGTPGACERGEQGPAGAPGFQGLPGPAGPPGEAG	645
Qy	1408	KPAHEGLVATVKAG--RSIHEIPREELRHTPELPLAPR-----PLKEGSITQGTPLK	1458
Db	646	KPGCEQVPGNAGAPGPAGARGERGFPGERGVOQPGPGQPGRGANGAPGNDGA-----	698
Qy	1459	YDTGASTTSGKKHDRVSLIGSGRTPPVHPPLDVMADARALERACYEESLKSRSPTASS	1518
Db	699	GDAGAP-----GAPNEGPP-----GLEHWPGBRGAA	725
Qy	1519	GGSIARGAPVIVPELGKPRQSPLTYDHGAPFAGHLPRGSPVPMREBFTPLRQEGSLSSK	1578
Db	726	GLPGKAG-----DRG-----DPGPKGADG-----	744
Qy	1579	ASQDKLTSTPREIAKSPHSTVPEHHPHPLISPYEHLRLGVSGVDLVRSHIPLAFDTSTIP	1638
Db	745	-----APKQGLRGLTG-----PIG--PPG-P	763
Qy	1639	RGIPLDAAAAYLPRHLNPYTPHLYPPYLIRGYPDTAALENRQTIINDYITSQQMHN	1698
Db	764	AGAPGDKGEG-----PPGPAGP-----TGA-----	784
Qy	1699	TATAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLQVPHLPVLVPPPTGTPATAMDRL	1758
Db	785	-----RG-----APDRG-----EPGPPGPA--GF	802
Qy	1759	AYLPTAP-QP-----PSSRHSSSPSPGGGTHLTKPTTTSSSRERDRDRDRDRE	1809
Db	803	AGPPGADQGPAGKGETGDAGAKGADGAPPGPAGPTGAPGA-----	842
Qy	1810	REKSILTSTTTVEHAPIWPPTGTSQSGSGSSGGGG--GSSSRPASHSHAHQHSIPSPRT	1867
Db	843	-----GZVGAPGPKAGSGAGPPGATGFPGAAGV-----GPPGPGSG	879
Qy	1868	ODALQORPSVLHNTKMG---IITAVEPSKPTVLRSTSTSSPVRPAATPPATHCPLGG	1923
Db	880	NIGLPGPPGPKGZGSKGPRGETGPAGRPCEP-----GPAGPPGPGEXKSGPCGADGPIGA	934
Qy	1924	TLDGVPYTLMEPVLVLPKEAPRVARPERPRADTCHAEFLAKPPARSGLF-PASSPSKSEPR	1982
Db	935	-----PGTGPQGIAGQGVGLPQO--RGERGFPLGPGPSGEPGKGQSGSASGERGPPG	987
Qy	1983	PLVPVPSGSHATIARTAKNLAPHASPDPPAPPA--SASDPHREKTKQSPFISIQELELRS	2040
Db	988	PMGPP-----GLACPPCE--AGREGAPGAEAGPCRDGAAGPKGDRGETGP-----	1030
Qy	2041	LGYHGSYSYPEGVEPVSPSSPLTHDKGLPKHLELDKSHLEGEURPK-----QPQGVKL	2096
Db	1031	AGPPGAPGAPGAPGVPVPAG-----KNGDRGETGPAGPAGPPGFPAGA	1072

DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(V) chain precursor.
GN COL5A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 556-565.
RX MEDLINE=91302336; PubMed=2071595;
RA Takahara K., Seto Y., Okasawa K., Okamoto N., Noda A., Yaoi Y.,
RA Kato I.;
RT "Complete primary structure of human collagen alpha 1 (V) chain.";
RL J. Biol. Chem. 266:13124-13129(1991).
RN [2]
RP SEQUENCE OF 621-822.
RC TISSUE=Chorioamniotic membrane;
RX MEDLINE=89227189; PubMed=2496661;
RA Sever J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of three
cyanogen bromide-derived peptides from human alpha 1(V) collagen
chain.";
RL Arch. Biochem. Biophys. 271:120-129(1989).
RN [3]
RP SEQUENCE OF 823-950, AND HEPARIN-BINDING.
RX MEDLINE=90366601; PubMed=2203476;
RA Yaoi Y., Hashimoto K., Koitabashi H., Takahara K., Ito M., Kato I.;
RT "Primary structure of the heparin-binding site of type V collagen.";
RL Biochim. Biophys. Acta 1035:139-145(1990).
RN [4]
RP SEQUENCE OF 556-571.
RC TISSUE=Placenta;
RX MEDLINE=92239022; PubMed=1571108;
RA Mann K.;
RT "Isolation of the alpha 3-chain of human type V collagen and
characterization by partial sequencing.";
RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).
RN [5]
RP SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.
RC TISSUE=Chorioamniotic membrane;
RX MEDLINE=94237164; PubMed=8181482;
RA Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champlaud M.F.,
RA Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT "Diversity in the processing events at the N-terminus of type-V
collagen.";
RL Eur. J. Biochem. 221:987-995(1994).
RN [6]
RP DISEASE, AND VARIANT EDS-I SER-1639.
RX MEDLINE=97195540; PubMed=9042913;
RA de Paep A., Nuytinc L., Hauser I., Anton-Lamprecht I.,
RA Naeyaert J.M.;
RT "Mutations in the COL5A1 gene are causal in the Ehlers-Danlos
syndromes I and II.";
RL Am. J. Hum. Genet. 60:547-554(1997).
RN [7]
RP VARIANTS EDS-1 SER-530 AND ASP-1489.
RX MEDLINE=20069401; PubMed=10602121;
RA Giunta C., Steinmann B.;
RT "Compound heterozygosity for a disease-causing G1489D and
disease-modifying G308 substitution in COL5A1 of a patient with the
classical type of Ehlers-Danlos syndrome: an explanation of
intrafamilial variability?";
RL Am. J. Med. Genet. 90:72-79(2000).
CC -!- FUNCTION: Type V collagen is a member of group I collagen
component of nearly ubiquitous distribution. Type V collagen binds
to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
CC -!- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
most tissues and trimers of one alpha 1(V), one alpha 2(V), and
one alpha 3(V) chains in placenta.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- PTM: Sulfated on 40% of tyrosines.
CC -!- DISEASE: Defects in COL5A1 are a cause of Ehlers-Danlos syndrome
type I (EDS-I) [MIM:130000]; also known as Ehlers-Danlos syndrome
gravis. EDS-I is a connective-tissue disorder characterized by
loose-jointedness and fragile, velvety, stretchable, bruisable
skin that heals with peculiar 'cigarette-paper' scars. Inheritance
is autosomal dominant.
CC -!- DISEASE: Defects in COL5A1 are a cause of Ehlers-Danlos syndrome
type II (EDS-II) [MIM:130010]; also known as Ehlers-Danlos
syndrome mitis. Inheritance is autosomal dominant.
CC -!- SIMILARITY: HIGH, TO ALPHA 3(V) AND ALPHA 1(XI) CHAINS.
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90279; BAA14323.1; -;
CC PDB: 1A89; 18-NOV-98.
CC PDB: 1A9A; 18-NOV-98.
CC Genew: HGNC:2209; COL5A1.
CC MIM: 120215; -;
CC MIM: 130000; -;
CC MIM: 130010; -;
CC GO: GO:0005588; C:collagen type V; TAS.
CC InterPro: IPR008161; Clg_helix.
CC InterPro: IPR008160; Collagen.
CC InterPro: IPR008985; ConA_like lec.gl.
CC InterPro: IPR000885; Fib_collagen_C.
CC InterPro: IPR001791; Laminin_G.
CC InterPro: IPR003129; TSPN.
CC Pfam: PF01410; COLFI; 1.
CC Pfam: PF02210; TSPN; 1.
CC ProDom: PD000007; Clg_helix; 1.
CC ProDom: PD002078; Fib_collagen_C; 1.
CC SMART: SM00038; COLFI; 1.
CC SMART: SM00282; LamG; 1.
CC SMART: SM00210; TSPN; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Collagen; Signal; Heparin-binding; Sulfation; Disease mutation;
CC Ehlers-Danlos syndrome; 3D-structure.
CC SIGNAL 1 37
CC CHAIN 38 1605
CC DOMAIN 39 230
CC DOMAIN 231 443
CC DOMAIN 444 558
CC DOMAIN 559 1570
CC DOMAIN 1571 1605
CC PROPEP 1606 1838
CC MOD_RES 234 234
CC MOD_RES 236 236
CC MOD_RES 240 240
CC MOD_RES 262 262
CC MOD_RES 263 263
CC MOD_RES 338 338
CC MOD_RES 340 340
CC MOD_RES 346 346
CC MOD_RES 347 347
CC MOD_RES 416 416
CC MOD_RES 417 417
CC MOD_RES 420 420
CC MOD_RES 421 421
CC MOD_RES 570 570
CC MOD_RES 576 576
CC MOD_RES 621 621
CC MOD_RES 627 627
CC MOD_RES 639 639
CC MOD_RES 642 642
CC COLLAGEN ALPHA 1(V) CHAIN.
CC TSP N-TERMINAL.
CC NONHELICAL REGION.
CC INTERRUPTED COLLAGENOUS REGION.
CC TRIPLE-HELICAL REGION.
CC NONHELICAL REGION.
CC CARBOXYL-TERMINAL PROPEPTIDE.
CC SULFATION (POTENTIAL).
CC SULFATION (POTENTIAL).
CC SULFATION (POTENTIAL).
CC SULFATION (POTENTIAL).
CC SULFATION (POTENTIAL).
CC SULFATION (POTENTIAL).
CC SULFATION (POTENTIAL).
CC SULFATION (POTENTIAL).
CC SULFATION (POTENTIAL).
CC HYDROXYLATION.
CC HYDROXYLATION.
CC HYDROXYLATION.
CC HYDROXYLATION.

FT MOD_RES 648 HYDROXYLATION. 648 2.7%; Score 352; DB 1; Length 1838;
FT MOD_RES 654 HYDROXYLATION. 654 Best Local Similarity 20.5%; Pred. No. 0.00014;
FT MOD_RES 657 HYDROXYLATION. 657 Matches 375; Conservative 104; Mismatches 595; Indels 756; Gaps 90;
FT MOD_RES 675 HYDROXYLATION. 675 842 EOKPPAAELAVDTGKAEPEVKSE-CTBEAEEG-----PAKGD 879
FT MOD_RES 678 HYDROXYLATION. 678 238 EHYSPDCDVTAPDTPSQSDPNDEYTYEGDEGETYYEYYPEDLDKEPTSK-KP 296
FT MOD_RES 680 HYDROXYLATION. 680 880 ABAABATAGALKAEKKEGGGRATTAKSGAPQSDSSATCSADEVDEAGDKNLLS 939
FT MOD_RES 686 HYDROXYLATION. 686 297 VEAAKETTE-----VPEELTPTTEAAPMPETSEGAKEDVG 334
FT MOD_RES 690 HYDROXYLATION. 690 940 -----PRPSLLTPT-----GDPRANASQKPLDLKQLKQRAAATPPQVTKVHEPP 985
FT MOD_RES 696 HYDROXYLATION. 696 335 IGDYDYVFESEYTPSPYDLDITYGEEN--PDQPTD-----PGAGA-----EIP 377
FT MOD_RES 705 HYDROXYLATION. 705 986 REDAAPTAPAPPPPO-----NLOPE-----SDAQPGSGSRKSRSPAPP 1029
FT MOD_RES 708 HYDROXYLATION. 708 378 TSTADTNSNPAPPPGGADLDGEFTTEIRNLDENYDYPDYDTSSP--SEIGOMP 435
FT MOD_RES 717 HYDROXYLATION. 717 1030 ADKEAF-----AAEAQKLGPPCWTGSLPFPVPPREVIKASPHAPDPSAFSAPP 1081
FT MOD_RES 720 HYDROXYLATION. 720
FT MOD_RES 726 HYDROXYLATION. 726
FT MOD_RES 732 HYDROXYLATION. 732
FT MOD_RES 744 HYDROXYLATION. 744
FT MOD_RES 756 HYDROXYLATION. 756
FT MOD_RES 762 HYDROXYLATION. 762
FT MOD_RES 765 HYDROXYLATION. 765
FT MOD_RES 771 HYDROXYLATION. 771
FT MOD_RES 774 HYDROXYLATION. 774
FT MOD_RES 780 HYDROXYLATION. 780
FT MOD_RES 789 HYDROXYLATION. 789
FT MOD_RES 795 HYDROXYLATION. 795
FT MOD_RES 804 HYDROXYLATION. 804
FT MOD_RES 807 HYDROXYLATION. 807
FT MOD_RES 810 HYDROXYLATION. 810
FT MOD_RES 816 HYDROXYLATION. 816
FT MOD_RES 819 HYDROXYLATION. 819
FT MOD_RES 834 HYDROXYLATION. 834
FT MOD_RES 846 HYDROXYLATION. 846
FT MOD_RES 861 HYDROXYLATION. 861
FT MOD_RES 864 HYDROXYLATION. 864
FT MOD_RES 870 HYDROXYLATION. 870
FT MOD_RES 873 HYDROXYLATION. 873
FT MOD_RES 876 HYDROXYLATION. 876
FT MOD_RES 882 HYDROXYLATION. 882
FT MOD_RES 888 HYDROXYLATION. 888
FT MOD_RES 891 HYDROXYLATION. 891
FT MOD_RES 897 HYDROXYLATION. 897
FT MOD_RES 903 HYDROXYLATION. 903
FT MOD_RES 906 HYDROXYLATION. 906
FT MOD_RES 930 HYDROXYLATION. 930
FT MOD_RES 945 HYDROXYLATION. 945
FT MOD_RES 1017 HYDROXYLATION. 1017
FT MOD_RES 1020 HYDROXYLATION. 1020
FT MOD_RES 1023 HYDROXYLATION. 1023
FT MOD_RES 1029 HYDROXYLATION. 1029
FT MOD_RES 1221 HYDROXYLATION. 1221
FT MOD_RES 1224 HYDROXYLATION. 1224

Db 436 ANODTIYEGIGPRGEKQK--GEPAITEPQMLIEGPP-----GPEGAGLPGPPG 484
Qy 1082 HPLPLG-----LHDTARPVLP RPPTTSNPPPLI-----SSAKHP--SVLER 1120
Db 485 TWGPTGVQDGERGPPRPGLPAGADGLPGPGTWMMLPFRFGGGDAGSKGPMWSAQES 544
Qy 1121 QIGAISQMSVOLHVYPYSEHAKAPVGPVTM--GLPLPMDPKKLAPFS--GVKQE-----QLS 1173
Db 545 QAAILQOARLAL-----RGPAGPMGLTGRPGVGP-----PGSGGLKGEPPGVGPG 592
Qy 1174 PRGAGPPESLGVPTAQEASVLRGTALGSVPGGSITKGIPISTRVPSDSAIYRGSIHTGT 1233
Db 593 PRGVQGPFPAGKPR-----RCRA-----GSDGARGMPGQTGP----- 626
Qy 1234 PADVLYKGTITRIIGEDSPSLDRGRED--SLPKGHVIVYEGKKHVLSEVEGMSVTQCSK 1291
Db 627 -----KG-----DRGFDGLAGLP-----GEKGI----- 644
Qy 1292 EDGRSSQPPHETAAPKRTYDMMEGRVG--RAI--SSASIEGLMGRAIPPERHSPHHLKEQH 1349
Db 645 RGDPGSGPPPGPGDDGERGD--DGEVGRGLPGKPGPRGLLPGKPGPPGPGP-----P 696
Qy 1350 HIRGSITQIGIPRSYVEAQEDYLREAKLLKREGTTPPPPPSRDLTEAYTQALGLKLP 1409
Db 697 GVTGMDGQPGKGNVGPQ-----GEFPGPGQGN----- 726
Qy 1410 AHEGLVATVKEAGRSIHEI PREELRHTPELPLAPLKEGSI TOGTPLKYDTGASTTGSK 1469
Db 727 GAQGLPG-----PQAGIP--PGKGLP--GKP----- 750
Qy 1470 KHDVRSLIGSPRTFPVPHPLDVNADALERACVYESLKRPGTASSSGSIARGAPVI 1529
Db 751 -----GLPGMGADGPGHP-----GKEGPPGEKG--QGPP-- 780
Qy 1530 VPGLKPRQSPPLYEDHGAPFAGHLPRGSPVTMBREPTPRLOEGLSSSKASQDRKLTSTP 1589
Db 781 -----GPQGPY-----PGPRGVK-----ADGIRGLKGTK 807
Qy 1590 REIAKSPHSTVPEHHPHIPSYEHLRGVSDVLYRSHIPLAFDPTSI-----PRGI----- 1641
Db 808 GE-----KGEDGPGFGKDMGIGKDRGETGPGPGEDGPE 843
Qy 1642 -PLDAAAAYLPHRLAPNPTYPHLYPPYLRGYDPTAALENRQTIINDYITSQMHNTA 1700
Db 844 GPKRGSGPGNDPGPLGPGKEKGLVPCL--PGY-----GRQ----- 879
Qy 1701 TAMAQRADMLKGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLPPTTGTATAMDRLAY 1760
Db 880 -----GPKGSGIF-----PGFPGANGKGR--G 901
Qy 1761 LPTAPQPPSSRSHSSPLSPGPGPHLTKEPTTSSSERERDRDRDREREKSIILTSITT 1820
Db 902 TPGKPRGRQGTGTPRGERGPRGIT----- 927
Qy 1821 VEHAPIWRPTEQSSGSSGSSG--GGGSSSRPASHAHQHSPTSPRTQALQORPSVL 1878
Db 928 -----GKPGKNSGGDPAGPGERGPNCPQGTGPGKPGPGKDLGHPGQR 981
Qy 1879 HNTQMKGIITAVEPSKPTVLSTSTSSPVRPAATFPATHCPGLGTLDGVTPLMEPVLL 1938
Db 982 GETGFG-----KTGPPGPGVVGPGQ-----PTGET----- 1008
Qy 1939 PKEAPRVARPERPRADTGHAPLAKPPARSGLF-----PASSPSKSGSP----- 1981
Db 1009 -----GPMGERGHGPPGPGGEGQLPGLAGKEGTGKDCPGAGLPGKDPGFLURG 1057
Qy 1982 -----RPLVPVSGHATTIARTPAKNLAPHASPDPPAP-----PASASD-----PHREK 2025
Db 1058 FPGDRGLPGPVG-----ALGLKNEGPPGPPGAGSGPGRGAGAGIPIGPGRG 1108
Qy 2026 TQSKPFSIQELELRSLGSHGSSYSPGVPSVSSPSLTHDKGLPKHL-----EELD 2078

Db 1109 PQGPP-----GPAGEKAGPEKGPQGPAGRDGLQGPVGLPGPAGVPVGPGEQGD 1157
Qy 2079 KSHL-----EGELRPQKPGVKGGLGGAHLPHLRPLPESQPSRSSPLLTATARGVK 2127
Db 1158 KEIGEPQKQSGKDKGEQGP--PGTGPQG-----PTGQPGPSGA---DGEQGP 1203
Qy 2128 GHRVVVTLAQHISEVITDYTRHHPQQLSAPLAPLYSPFPGASCPLDLRPPPSDL----- 2183
Db 1204 GQGL--FGQKDE-----GPRGFP--GPPGVPVGLQGLPG-----PPGEKGTG 1243
Qy 2184 ----YLPDPDHGAPAGSHSEGGKSPENKTVSGGDEGIEPVSPPEGMTPEGHRSRA 2240
Db 1244 DVQMGPP-----GPPGPRGPGCAP-----GADG--PQGPFGIGNPG-----A 1280
Qy 2241 VYPLLYRQGEOTEPSEMSKSGPNTSQPARFSKLTESNAWVSKQEIKNLTHNRN 2300
Db 1281 V-----GEKEPGEAGEPGSGSGPP-----GPKGERGEK-----GES 1314
Qy 2301 EPEYNISQGTBIFNMPAITGTGLMYRSQAVOEHAHSTNMGLEAIRKALMGKY-----DQ 2356
Db 1315 GFSGAAGPGRPK--GPPGDDGP-----KGSFGPVGFPDGP 1347
Qy 2357 WEESPPLSNAFNPPLNASASLPAAMPITADGRSDHTLTSPGGGKAKVSGRPSRKAKS 2416
Db 1348 GPPGEPGAGQDGPFGDKGD-----DGEPGQT--GSPGTPGEPGPGPKRGP 1396
Qy 2417 PA-PGLASGDRPSPVSVHSEGDNRRTPLTNRVWEDRPSSAG--STPPFYNPLIMRLOA 2473
Db 1397 PGPGRQGEK--GAKGEAGLEPGPKTGIPQGAAPGPGDGLRGIPGPVG-----EQ 1449
Qy 2474 GVMASPPPPGLPAGSGPLAGPHAWDEEPK 2503
Db 1450 GLPGSGFGDPPGPGWPPGLPGLKGDGPK 1479

RESULT 53

CA2B HUMAN
ID CA2B HUMAN STANDARD; PRT; 1736 AA.
AC P13942; Q07751; Q13271; Q13272; Q13273; Q99866; Q9UIP9;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 2(XI) chain precursor.
GN COL11A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96032717; PubMed=7559422;
RA Vuorio M.M., Pihlajamaa T., Vandenberg P., Prockop D.J.,
RA Ala-Kokko L.;
RT "The human COL11A2 gene structure indicates that the gene has not
RT evolved with the genes for the major fibrillar collagens.";
RL J. Biol. Chem. 270:22873-22881(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Tubby B.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 59-807 FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93314796; PubMed=8325374;
RA Zhidkova N.I., Brewton R.G., Wayne R.;
RT "Molecular cloning of PAPP (proline/arginine-rich protein) from human
RT cartilage and subsequent demonstration that PAPP is a fragment of the
RT NH2-terminal domain of the collagen alpha 2(XI) chain.";
RL FEBS Lett. 326:25-28(1993).
RN [4]
RP SEQUENCE OF 730-1690 FROM N.A.
RX MEDLINE=89340485; PubMed=2760050;
RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,

van der Rest M., Ono K., Solomon E., Ninomiya Y., Olsen B.R.;
RT "The human alpha 2(XI) collagen (COL11A2) chain. Molecular cloning of
RT cDNA and genomic DNA reveals characteristics of a fibrillar collagen
RT with differences in genomic organization.";
RL J. Biol. Chem. 264:13910-13916(1989).
RN [5]
RP SEQUENCE OF 1-537 FROM N.A.
RX MEDLINE=96435918; PubMed=8839804;
RA Lui V.C., Ng L.J., Sat E.W., Cheah K.S.;
RT "The human alpha 2(XI) collagen gene (COL11A2): completion of coding
RT information, identification of the promoter sequence, and precise
RT localization within the major histocompatibility complex reveal
RT overlap with the KX5 gene.";
RL Genomics 32:401-412(1996).
RN [6]
RP ALTERNATIVE SPLICING.
RX MEDLINE=95239468; PubMed=7721876;
RA Zhidkova N.I., Justice S.K., Wayne R.;
RT "Alternative mRNA processing occurs in the variable region of the
RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";
RL J. Biol. Chem. 270:9486-9493(1995).
RN [7]
RP DISEASE.
RX MEDLINE=20143361; PubMed=10677296;
RA Melkonian M., Brunner H.G., Manouvrier S., Hennekam R.,
RA Superti-Furga A., Kaeerlaeinen H., Pauli R.M., van Essen T.,
RA Warman M.L., Bonaventura J., Miny P., Ala-Kokko L.;
RT "Autosomal recessive disorder otospondyloymegeophysal dysplasia is
RT associated with loss-of-function mutations in the COL11A2 gene.";
RL Am. J. Hum. Genet. 66:368-377(2000).
RN [8]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [9]
RP VARIANT OSMED ARG-661.
RX MEDLINE=95163096; PubMed=7859284;
RA Viskula M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,
RA Goldring M.B., van Beersum S.E.C., de Waal Malefijt M.C.,
RA van den Hoogen F.H.J., Ropers H.-H., Wayne R., Cheah K.S.E.,
RA Olsen B.R., Warman M.L., Brunner H.G.;
RT "Autosomal dominant and recessive osteochondrodysplasias associated
RT with the COL11A2 locus.";
RL Cell 80:431-437(1995).
RN [10]
RP VARIANTS GLY-593; LYS-824; LEU-879; THR-1316 AND GLN-1600.
RX MEDLINE=98254467; PubMed=9585596;
RA Koga H., Sakou T., Taketomi E., Hayashi K., Numasawa T., Harata S.,
RA Yone K., Matsunaga S., Otterud B., Inoue I., Leppert M.;
RT "Genetic mapping of ossification of the posterior longitudinal
RT ligament of the spine.";
RL Am. J. Hum. Genet. 62:1460-1467(1998).
RN [11]
RP VARIANT WZS GLU-955.
RX MEDLINE=99021942; PubMed=9805126;
RA Pihlajamaa T., Prockop D.J., Faber J., Winterpacht A., Zabel B.,
RA Gledhill A., Wiesbauer P., Spranger J., Ala-Kokko L.;
RT "Heterozygous glycine substitution in the COL11A2 gene in the original
RT patient with the Weissenbacher-Zweymueller syndrome demonstrates its
RT identity with heterozygous OSMED (nonocular Stickler syndrome).";
RL Am. J. Med. Genet. 80:115-120(1998).
RN [12]
RP VARIANT STL3 940-GLY--PRO-948 DEL.
RX MEDLINE=98165506; PubMed=9506662;
RA Sirko-Osada D.A., Murray M.A., Scott J.A., Lavery M.A., Warman M.L.,
RA Robin N.H.;
RT "Stickler syndrome without eye involvement is caused by mutations in
RT COL11A2, the gene encoding the alpha-2(XI) chain of type XI
RT collagen.";

Db 692 HPGKEGPGCT-----KGNQSPGPGQGLVGYGP-RGVKGVGDGIRGLKHGKEG-EDGPPG 745
Qy 1265 -KGHVIVECKKHVLSYEGCHSVTQCSKEDG-----RSSGPHETAAP-----KRTYD 1312
Db 746 FKXGDIIVKGRGEV-----GVPSRGEDGPEGKGRGTGTPDPPGLMGXGKGLGP 798
Qy 1313 MMEGRVGR-----AISSASTEGLMGRAPPERHSPHHLKEQHIGSI--- 1355
Db 799 GLPGYPGQKPGKSLGPFPGCASKGKAGLSGKSGRGERGTPGPRGQGRPGATGKS 858
Qy 1356 -----TGICPRSYVEAQEDYLRREAKLLKREGTP-----PPPPFSRDUTE 1395
Db 859 GAKGTSGDGHGPPGKGLPGPQ-----GPNRPGKPGKPGPGKGLP 903
Qy 1396 AYKTO-----ALGPLKLPAHGLVATVKEAGRSIHEIPRELRTHTPELPLAPRLK 1447
Db 904 GHPGQGEVFGKTKGP-----PGPGVVGPOQAAGET---GPMGERGH---PGPPGPG 952
Qy 1448 EGSITQGTPLKYDTGASTTGSKKHVRSLIGSPG-----RTFPVPHPLDVMADAR 1497
Db 953 E-----QGLP-----GTAG--KEGTDGDPGPGAGPKDGPAGLGFPGERGLGTAGGP 999
Qy 1498 ALERACYEESLKSRTGTASSG---GSIARGAPVIVPELKGKPRQSPLYTHEDHGAPFAGHLP 1555
Db 1000 GLKG---NEGSPGPPGAGSPGERGAGSGGPIGPP---GRPG-----P 1037
Qy 1556 RGSPTVMEPTPRLOEGSLSSKASQDKLSTPREIAKSPHSTVPEHHPIPSYEHLL 1615
Db 1038 QGPPGAAGEKGVPGKGPIG-----FTGRDVGQGVGL-----PGFAGP--- 1076
Qy 1616 RGVSGVDLYRSHIPLAFDPTSPRIGIPLDAAAAYVLPRLAPNPTYPHLYPYLIRGYPD 1675
Db 1077 PGVAGEDGKGEV-----GDPQKGTGKNKGEGHPGPGP-IGPV-----GQPG 1120
Qy 1676 TAALENRTIINDYITSOQMHNTATAMAQADMLRGLSPRESSALNAYAGPRGIIDL 1735
Db 1121 AAGADGEPG-----ARGPQGHFG-----AKDEGTRGFNGPGPIGLQGLPGPSG--EKG 1168
Qy 1736 QVPHLPVLVPTPGTPTAMDRLAVLTPAPQFSSRHSLSLP-----GGTHLTKP 1798
Db 1169 ETGDVGMPGPGP-----PGPRFAGNGADGQGGPGGVGNLGPPEGKEP 1215
Qy 1789 TTTSSSERDRDRDREREKSIILSTTTTVEHAPIWRGTQSSGSSSGSGG--G 1846
Db 1216 GESGSPGIQGEVPGKPRGERGEG--ESGQGEPPG---PGAKGPGQDDGPKGNPVPV 1270
Qy 1847 SSSRPASHAHQHSPISPRTQDALQORPSVLHNTGKGIITAVEPSKPTVLRSTSSP 1906
Db 1271 FPGDPPGPEG-----GPRGQDGAK-----GDRGEDG-----EPGQP-----GSP 1305
Qy 1907 VRPAATRPATHCPGLGTLGQVYPLMEPVLLPKAEPRVARPERPRADTGHAFKAPAR 1966
Db 1306 GPTGEPGPG---PLG-----PLG-----PLG-----PLG-----K 1319
Qy 1967 SGLFPASSPSKSGEPRLPVVPSGHATTIARTPAKNLAPHASDPDPAPPASADPHREKT 2026
Db 1320 RG--PAGSP--GSEGR-----QGKGAKGDPGALGAPKGTGPVGPAGPAGKPGP--DGL 1367
Qy 2027 QSKPFSI-QEELSLSLGHSYSPEGVEPVSPVSSPSLTHDKGLPKHLELDKSH--LE 2083
Db 1368 RGLPGSVQQGRPGATGQAG-----PPG--PVGPFLGLRGLRDAG-----AKGKGHFGLI 1416
Qy 2084 GELRPKQPGVKLGEEAHLPHRLPESQSSPSPLLTQAPGVKGHQRVVTLAQHSI SEVI 2143
Db 1417 CLIGP--PGEQGEKD-----RGLPGQGS-----PGQGEKGI----- 1448
Qy 2144 TDYTRHHPOQLSAPLAPLYSFFCASCPCVLDLRRPPLDLYPPPDHGCAPGSPHSEGG 2203
Db 1449 -----PGASGPI-----GPGQ-----PGP-----GLPGPAGPKGAKG 1474
Qy 2204 KESPEPNKTSVLGGEDGIEVSPGEMTGPCHSRSAVYPLLYDDGQGTPEPSRMGSK--- 2260
Db 1475 ATGP-----GGPKGEGVQGPFG--HFGPPGEVIQPLFIQMPKTRRSVDGSLMQ 1523

Qy 2261 -----SPGNTSQPPAPFSPKLTESNANVKSQKOEINK 2292
Db 1524 EDEAIPGTGAPGSPGLEIFGSL-----DSLREEIEQ 1556
RESULT 54
APC_HUMAN
ID APC_HUMAN STANDARD; PRT; 2843 AA.
AC P25054; Q15162; Q15163;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein).
GN APC OR DP2.5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN
SEQUENCE FROM N.A.
RX MEDLINE=91335210; PubMed=1651562;
RA Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,
RA Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,
RA Finnear R., Markham A., Groffen J., Boguski M.S., Altschul S.P.,
RA Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.,
RT "Identification of FAP locus genes from chromosome 5q21.";
RL Science 253:661-665 (1991).
[2]
RN
SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RP TISSUE=Fetal brain;
RX MEDLINE=91330307; PubMed=1678319;
RA Joslyn G., Carlson J., Thliveris A., Albertsen H., Gelbert L.,
RA Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,
RA Sargent L., Krapcho K., Wolff E., Burt R., Hughes J.P.,
RA Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,
RA Abderrahim H., Cohen D., Leppert M., White R.,
RT "Identification of deletion mutations and three new genes at the
familial polyposis locus.";
RL Cell 66:601-613 (1991).
[3]
RN
ASSOCIATION WITH CATENINS.
RP MEDLINE=94082295; PubMed=8259519;
RA Su L.-K., Vogelstein B., Kinzler K.W.;
RT "Association of the APC tumor suppressor protein with catenins.";
RL Science 262:1734-1737 (1993).
[4]
RN
DISEASE.
RP MEDLINE=95174843; PubMed=7661930;
RA Hamilton S.R., Liu B., Parsons R.E., Papadopoulos N., Jen J.,
RA Powell S.M., Krush A.J., Berk T., Cohen Z., Tetu B., Burger P.C.,
RA Wood P.A., Taqi F., Booker S.V., Petersen G.M., Offerhaus G.J.A.,
RA Tarmette A.C., Giardiello F.M., Vogelstein B., Kinzler K.W.;
RT "The molecular basis of Turcot's syndrome.";
RL New Engl. J. Med. 332:839-847 (1995).
[5]
RN
DISEASE.
RP MEDLINE=97094176; PubMed=8940264;
RA Eccles D.M., van der Luijt R.B., Breukel C., Bullman H., Bunyan D.,
RA Fisher A., Barber J., du Boulay C., Primrose J., Burn J., Fodde R.;
RT "Hereditary desmoid disease due to a frameshift mutation at codon 1924
of the APC gene.";
RL Am. J. Hum. Genet. 59:1193-1201 (1996).
[6]
RN
DISEASE.
RP MEDLINE=20243021; PubMed=10782927;
RA Couture J., Mitri A., Lagace R., Smits R., Berk T., Bouchard H.-L.,
RA Fodde R., Alman B., Bapat B.;
RT "A germline mutation at the extreme 3-prime end of the APC gene
results in a severe desmoid phenotype and is associated with
overexpression of beta-catenin in the desmoid tumor.";
RL Clin. Genet. 57:205-212 (2000).
[7]
RN

QY 2096 LGG--EAAHLPHLRPLSPESQSSSPLLOTAPGVKGQRVWTLAQHISEVITQDYTRHHQ 2153
Db 2543 RSGTWKREHSKH-----SSSLPRVSTWRTTSSSSILSASSESEKAKSEDEKH--- 2591
QY 2154 QLSAPLAPLYSPFGASCPLVLDLRPPSDLYLPPDPHCAPARGSPHSEGGKRSPEPNKTS 2213
Db 2592 -----VNSISGTK-----QSKENQVSAGTWKRIKENEFPSTNSTS 2627
QY 2214 --VLGGGEGDIE-----PVSPPEGMTEPHGHSRAVYPLLYRDGEGTEPFRMGSKSPGNT 2265
Db 2628 QTVSSGATNGAESKTLIYQMAPAVSKTEDVWRIEDCP-----NNPRSGRSPGNT 2679
QY 2266 SQPAFFSKLTESNAMYKSKQEKINKLTHNRNEPYNISQPTETFPNPAITGTGLM 2325
Db 2680 --PP-----VDSVSEKANPNIKOSKONQAKQV-----GNGSV 2711
QY 2326 TYRQAVQEHASTNMGLEAIRKALMGKYDOWEESPPLSANAFNPLNASISLPAAMPITA 2385
Db 2712 PMRTVGLNRLNLSFIQVADPOQKTEIKPGQNNPVVSETN-----ESSIVERTPPSS 2764
QY 2386 ADGRSDHTLTSPGGGKAKVSG--RPSRSKAKSPAGLASGDRPPSVSVHSEGDGNCNR 2442
Db 2765 SSS-SKH--SSPSGTVARVTPFNPNPSRKSSAD-----STSARPSQIP----- 2806
QY 2443 TPLNTRVWEDRPSAGST 2460
Db 2807 TPVANNNT-KGRDSKTGST 2823

RESULT 55

ZAP3 HUMAN
ID ZAP3 HUMAN STANDARD; PRT; 1822 AA.
AC P49750; P49752; Q9P1V7;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear protein ZAP3 (ZAP113).
GN ZAP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Baradarani L., Birditt B.,
RA Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,
RA Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.,
RT "Sequencing of human chromosome 14q24.3 region."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95319502; PubMed=7596406;
RA Sherrington R., Rogeev E.I., Liang Y., Rogeev E.A., Levesque G.,
RA Ikeda M., Chi H., Lin C., Li G., Holman K., Teuda T., Mar L.,
RA Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I.,
RA Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,
RA Saneau P., Pollinsky R.J., Wasco W., da Silva H.A.R., Haines J.L.,
RA Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
RA Rommens J.M., St George-Hyslop P.H.;
RT "Cloning of a gene bearing missense mutations in early-onset familial
RT Alzheimer's disease."
RL Nature 375:754-760(1995).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1661.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.
CC EMBL; AC007956; AAF61275.1; -
DR EMBL; L40403; AAC42008.1; ALT_FRAME.
DR EMBL; L40400; AAC42006.1; -
KW Nuclear protein.
FT DOMAIN 15 205 PRO-RICH.
FT DOMAIN 382 430 GLN-RICH.
FT DOMAIN 807 1209 ARG-RICH.
FT DOMAIN 1488 1577 ARG-RICH.
FT CONFLICT 621 621 P -> S (IN REF. 2).
FT CONFLICT 1404 1404 T -> I (IN REF. 2).
FT CONFLICT 1821 1821 K -> E (IN REF. 2).
SQ SEQUENCE 1822 AA; 204947 MW; 8E6CB83FE540C7D2 CRC64;
Query Match 2.6%; Score 349.5; DB 1; Length 1822;
Best Local Similarity 20.2%; Pred. No. 0.00016;
Matches 370; Conservative 157; Mismatches 642; Indels 663; Gaps 91;
QY 767 PKPATILG---ADGPPFPPTTP-----RRTSRAPTEPTPASEA 802
Db 87 PPPVMPGGYGDQNPMPMPPPPPGPPALSYKQKQYKHQMLHQDQDPPGLVPMELSP 146
QY 803 TGAPTPPPA---PSPSPAPPVVPKKEEETAA---APPVEEGEEQKPPAAEELAVDTG 856
Db 147 PESPPVPPGSMPP 202
QY 857 KAEPPVSECTEAEAGPKADAEAAEATAGALKAEKKEGGSGRATTAKSS----- 909
Db 203 ----PTPS-----YSSSSSSQSYLSHSQSYLSPSSQSPSRQSHSKSQ 243
QY 910 --GAPOQSDSATCSADEVDEAGGDKNRLISPRSLTTPGDPANASPKQLDLKQK 967
Db 244 LLAPPPSAPPNGKTTVQEPLESGAKNK--STEQQAAPEDP---STWTPQEQQYW 297
QY 968 QRAAAIPIQVTKVH-----EPPREDA-----APTAPAPPAPP- 1001
Db 298 YRQHLLSLQRTKVHLPGHKKGVPVAKDTPEPVKEEVTVPATSVPESSPEEPPLPPN 357
QY 1002 ---PQNLPQSDAPQQGSSPRGKSRSPA----- 1027
Db 358 EEPVPLPPEPQSEDEEDARLQLOAAAHQHQHVRVGFQVQGMKHTQLQILQ 417
QY 1028 -----PPADKEAFAAAQKLPDPPCWTSGLPFPVPPPREVIKASHPADPSAFSYAPP 1080
Db 418 QYQIIOPPPHIQATTP---PGIPP---PGVQGIPO--LTAAPVPASSQSSQVP 468
QY 1081 GHPLPLGLHDTARVLPPTISNPPPLISSAKHPSVLEROIKAISQMSVOLH----- 1134
Db 469 EKRP-----ALLPTVSVFGSAPP---TTHPPLQ-----SAGPSQVNSKAPLS 510
QY 1135 ---VPYS-----SHAKAPVGVMTGLPLMDPKLAP-----FSGV 1167
Db 511 KSALPVSFSSDQGLGESSAAPSQTITAVKMDVRSGLLPPDRSSVLESPPRPFDPG 570
QY 1168 KQ-EQL-----SPRGOAGPESLGVPTAQEASVLRGT--ALGSVPQGSITKG--I 1212
Db 571 RRFEDLSRCGPRPKGRFEGNEPD--GPRPRYEGHPAECTKSKWGMIPGPASQFVIT 628
QY 1213 PSTRV-PDSAITVRG----- 1230
Db 629 PSTSLSPRQSGPWKPKPAFGQHQKQPKSQABSLSGNKEPLADTSSNQKNFMQSA 688
QY 1231 HGTPADVLTKTITRIIGEDSPSLDRGREDLSKGHVIEYKKGHVLSYEGGMSV----- 1286
Db 689 FSTAADV-----KDVKAASNENLSDSQEP-PKSEV-----SEGPVPSNWDQVQSM 737
QY 1287 TOCKEDGRSSSG------PHETAAPKRTYDMMEGRVGRATSSASIEG-----LMGRAI 1335
Db 738 TQIDKAAVTPQVPLANKVPQAQSTFSKT--GGNEG--GTAVATSSLTADNDFPVGIGL 794
QY 1336 PPERHSPHLKEQHHRGSIITQGIIPRSYVEAQEDYLREAKLKRGTTPP-----PPPSR- 1391

Db 795 P-----HSENNQDKGLPRPD-----NRDNRLGNRGNSSSYRGPGQSRM 833
QY 1392 -----DLTEAVKTAAL--GPLKKAHEGLVATV-----KEAGRSIHRLPREELRHTPEL 1439
Db 834 ETRDKGLVNRGQAIISRG-----GLVKQEDFRDKWGR--REDSREKNR-----879
QY 1440 PLAPRPLKEGSIQTGTPKLYDTGASTTGKXHDV--RSLIGSPGRTFFPVPHVPLVMDARA 1498
Db 880 -----GEGS-----RUGLVRPGSSREKVPGLQSQDR-----908
QY 1499 LERACYEESLSRPTASSGSGSIARGAPVIVPELKGPRQSP-----TYEDHGAPFAGHL 1554
Db 909 -----CAAGSRERGGP-----RRAGSQERGRLRAGSRERIPRRAGSR 947
QY 1555 PRGSPVTREPTPRLOEGSLSSKASQDR--KLSTPREIAKSPHSTVPEHPPH--PISPYE 1612
Db 948 ERGPP-----RGQSR--EGLGRSDFRGRPFEPGDKGMY---PYHDEPFPPAPWN 999
QY 1613 HLLRGVSGVDLYRSHIPLAFTPSIP-----RGIPLDAAAAYLPR 1653
Db 1000 H-----GEE--RGHEEFLDGRNAPMERERLDDWDRERYWRECERDYQDDTLLEYNRED 1051
QY 1654 HLAENPTVPHLYPPVILRGYDPTALENRQTIINDYITSQMHNN---TATAMAQADML 1710
Db 1052 RFSAPPSRSHDGD-----RRGFWDDWDRDQMDDEDY--NREMERDMRDVDRISRPMY 1105
QY 1711 -----RGLSPRESSLALNYAAGRGIIDLQVPLVLVPTTGTATAMDR 1757
Db 1106 DRSLDNEWDYGRPLDQESQFR-----DIPSLPLPL--PPLPPLDRYDDR 1155
QY 1758 LAYLPTAQPFSSRHS--SSPLSPGPTHTLKTPTTSSSERR-----DRDRERDRDR 1808
Db 1156 WR-----EERNREGYDRDFRDELRIREYPERGDTWREKRDVYVDMWDERERLSDR 1209
QY 1809 EREKILSTTTVEHAPTWRCTEQ---SSGSGSGSGGGG---SSSPASHSHAHQSP 1863
Db 1210 WYPSDVRSHSPMAEHMPSHSSSEMMSGSDSLSDQGLGVVLSQR-----QHEII 1261
QY 1864 SPRTQDALQRPSPVLNHTGMGIITAVEPS-----KPTVLRTSTSSPVRPAATFPATH 1918
Db 1262 LKAQELKMLREKQELQKMDP--GSEFQWADHLPPQESLQNTSS--RCMTYPPGSGY 1317
QY 1919 CPLGCTLDGVYFTLMEPVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEPASP 1978
Db 1318 RP-----PPMCKPGSIVRP-----SAPPARSSV--PVTRPVP 1350
QY 1979 SEPRPLVPPV-----SCHATART 1997
Db 1351 IPPPPPPPPPPPPPPVVKPQTSVAEQERWDEDSFYGLWDTNDEQGLNSEFKSETAAIPSA 1410
QY 1998 PAKNLAPHASPDPPAPPASADPHREKTQSKPFSIQLELRLSLGY--HGSSYSPEGVFPV 2056
Db 1411 PVLPPPPVHSSIPPGCPVPMGMP-----MSKPPVQO-----TYDYGHRDSTNKVEOI 1461
QY 2057 -----SPVSSPSL-----TH 2066
Db 1462 PYGERITLRPDPLPERSFTFETHAGQDRYDRERDREPYFDRQSNVIADHRDFKDRDRETH 1521
QY 2067 -DKGLPHLELDKSHLEGELRPKQPGVKLGGEAHLPLHR---PIPESQSPSPLOT 2122
Db 1522 RDRDRGVIDYDRDRFRDRRDRDRAQSYRDKKDHSSRRGGFDRPSYDRKSDRFPVE 1581
QY 2123 AFGVKGHORVVTLAHQISEVITQDVTTRHHPOOLGAPLPAPLYSPFGASCPVLDLRRRPPS- 2181
Db 1582 GPSMFGGER-----RTYPE--RWPLPAPLSLHQPPAPRVE--KPKESK 1622
QY 2182 ---DLYLDP-----PD-----HGAPKSGSPH 2199
Db 1623 NVDDILKPPGSRSPRERIVWIMRGLPGSGKTH 1654

RESULT 56
CAL3_MOUSE

ID AC P08121: 061429: 09CEN7: STANDARD; PRT; 1464 AA.
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X DBA: TISSUE=Embryo;
RX MEDLINE=95011609; PubMed=7926795;
RA Toman D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
RT complete DNA sequence.";
RL Gene 147:161-168(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-488 FROM N.A.
RX MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogel G.;
RT "Complete nucleotide sequence of the N-terminal domains of the murine
RT alpha-1 type-III collagen chain.";
RL Gene 61:225-230(1987).
RN [4]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=8511189; PubMed=3972847;
RA Liau G., Mudryj M., de Crombrughe B.;
RT "Identification of the promoter and first exon of the mouse alpha 1
RT (III) collagen gene.";
RL J. Biol. Chem. 260:3773-3777(1985).
RN [5]
RP SEQUENCE OF 810-1464 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,


```
Db 713 GPFGASGSPQLQMPGCGP---GSPGKGKGPGGAGAD----- 751
Qy 1599 TYPEHHPHPISEYHLLRGVSDYLRSHIPLAFDPTSI PRGIPLDAAAAYLPHLAPN 1658
Db 752 -----GVPGKD-----GPRG----- 761
Qy 1659 PTYPHLYPPYLRGYPDTAALENQTIINDYITISQMHNTATAMAQADMLRGLSPRES 1718
Db 762 PAGP-IGPP-----GPAGQCDKGEGSP--- 784
Qy 1719 SIALNYAGPRGIIDLSQVPHLPVLVPTPGPATAMDRLAYLPTAPQFFSRHSSPLS 1778
Db 785 --GLPGIAGPRG-----CPG-----ERGEHGPAGPAGF----- 810
Qy 1779 PGCPHTLTPTTSSSERDRDRDRREKSIILTSITVEHAPITWRGTEOSSSS 1838
Db 811 PGAPONGEP-----GAKERGAPGKGEGPGGPGAGTGS-----GPAGPGPGQGVKGER 862
Qy 1839 GSSGGG-----GSSSRPASHAHQHSPISRTQDALQRPVSLHNTGMKGIITAVEPS 1893
Db 863 GSPGGGTAGTGGGRLCPGPNNGNPGPPGSGAPGKDGPPGAGNSG-----SPG 914
Qy 1894 KPTVLRSTSTSPV-----RPAATFPATHCPGLGTLGVYPTLMBPVLLPKPRVARPE 1949
Db 915 NPGIAGPKGDAGQPGKGPAGGPGGPGPLG--IAGL----- 951
Qy 1950 RPRADTGHAFKAPPARSGLEPASSPS-----KGSEPRDLPVPSVSHATARTPAKNLAPH 2005
Db 952 -----TGARGLAGPPGMPG--PRGSPGPGQKGSGKP-----GASGH 987
Qy 2006 HASPDPPAP-----PASASDPHREKTSQKPFISQELRLSLGVHSGSYSPGVEPVSPV 2059
Db 988 NGERGPGPGQGLPGQGTAGPGRDGNPG-----SDGQPRDGSPPGKGDRCEN 1036
Qy 2060 SPSLTHDKGPKHLELDKSHLEGELRPKQPPVKLGGEAAHPLHPLRPESQFSSP- 2118
Db 1037 GSPGA-----PGAPGH-----PGPPGVPGSKGDRGETGP---AGPSGAPG 1076
Qy 2119 --LLQTA-----GVKGHORVVTLAQHISEVITQDYTHHQQQLSAP 2158
Db 1077 PAGARGAPGPGPGKGTGERSNGIKGHR----- 1108
Qy 2159 LPAPLYSPFGASCPVLDLRRPSSDLYLPPPDHGAP-----ARGSPHSEGGKSPBPNTSV 2214
Db 1109 -----GFFG-----NPG-----PPGSPGAAGHGAIGSPGAPGPGVPHGPP- 1147
Qy 2215 LGGEDGTE-----PVSP-----EGWTEPHSRSAVYPLLRYDGEQTSPRWGS 2259
Db 1148 ---GKDGTSGHGPIGPPGPRGNRGERGSEG--SPGH-----PGQPGPPGPPGA 1191
Qy 2260 KSP-----GNTS-----QPPAFTSKLTESNAMSVMKSKQKQINKKLANTH- 2297
Db 1192 GPCCGGAAGVGGGSGSPFYGDDPMDPKINTEINSLKSVNGQIESLSPDG 1251
Qy 2298 NRNEPEYN 2305
Db 1252 SRKNPARN 1259

RESULT 57
CA14_MOUSE
ID CA14_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DC 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
```

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RL J. Biol. Chem. 264:6310-6317(1989).
RN RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RL sequence of the alpha 1 chain of basement membrane (type IV) collagen
FEBS Lett. 227:5-8(1988).
RN RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RL synthetic oligodeoxynucleotide."
Gene 43:301-304(1986).
RN RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RL Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RL the alpha 1(IV) chain of basement membrane collagen as derived from
RL complementary DNA."
Eur. J. Biochem. 147:217-224(1985).
RN RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RL Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RL alpha 1(IV) and alpha 2(IV) collagen."
J. Biol. Chem. 262:8496-8499(1987).
RN RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RL "Alpha 1 type IV collagen gene evolved differently from fibrillar
RL collagen genes."
J. Biol. Chem. 261:6654-6657(1986).
RN RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RL "Head-to-head arrangement of murine type IV collagen genes."
J. Biol. Chem. 263:19274-19277(1988).
RN RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burbelo P.D., Martin G.R., Yamada Y.;
RL "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RL bidirectional promoter and a shared enhancer."
Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;
RL "Structure of the amino-terminal portion of the murine alpha 1(IV)
RL collagen chain and the corresponding region of the gene."
J. Biol. Chem. 263:8706-8709(1988).
CC -/- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -/- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
```


Db 777 -----LQIRG-----DEG--PPGVQGPAGP----- 795
Qy 1658 NPTYPHLYPPYLI-----RGYPTAALNRQTIINDYIITSQQMHH 1697
Db 796 -PGVPGIGPPGAMGPPGEGPPGSGPPGIGKEGFGPGGLD----- 837
Qy 1698 NTATAMAQRADMLRGLSPRESSLALNVAAGPRGIIIDLSQVPHLPVL--VPPTPGTPATAM 1755
Db 838 -----MPGPKDKSQGLPGLTGOSGL-----PGLPGQQTGPGVPFGP-SK 878
Qy 1756 DRLAYLTPAOPFSSRHSSPLSPGCPHLYTKPTTSSSERDRDRDRDRERREKSIL 1815
Db 879 GEMGVMTGPGPGSPGAGTGLPG-----ERK-- 906
Qy 1816 TSTTVEHAPLWRCTEQQSSGSSGGG--GSSSRPASHAHQHSPIPRQDALQQ 1873
Db 907 -----DHG-----LPGSSGFRGDPGKDGKDVGLPGMPSMEHDMG---SMKQKGDQG 954
Qy 1874 RPSVLHNTKMGIIITAVEPSKPTVLRSTSTSSPVPPAATFPATHCPGLGTLGTVYPTLM 1933
Db 955 EKGQIGPTGDK-----SRGDPGTPG----- 975
Qy 1934 EPVLLPKAPVARPER--PRADTGHAFLAKPPARSGL-----PASSPSKGE 1980
Db 976 ---VPGKDGQAGHPGQPGPKGDRG---LSGTPGSGPLPGKSGVGMGLPGSGPEKG-- 1026
Qy 1981 PRPLVPPVSGHATTARTPAKNLAPHASPDPPAPASADPHREKTSKPSIQELESLS 2040
Db 1027 ---VPGIPGSGVPGSGEKA-----KGEKQG----- 1052
Qy 2041 LGYHGSSYSPGFBFVSPVSSPSLTHDKGL-----PKHLEELDKSHLEGELRPKQGPV 2094
Db 1053 -----GLPGIGIPGRPGDKDGLAGFPGSGEKEKGSAGTPG--MPGSPGR 1099
Qy 2095 KLGEAAHLPHRLPL-ESQSSSPLQTAGVKGHORVTLAHI SEVITQDYTRHHQP 2153
Db 1100 GSPGNIGH-PSGSLPGKDGKGLPLGDLGVPGVRG----- 1133
Qy 2154 QLSAPLAPLVSFFCASFVLDRRPPSDLYLP-----PPDHGAPARGSPHSGGKRSPE 2208
Db 1134 --EAGLPQT---PGTPGAGKGEPSGD-GIPGSAKEKGQGVPGKGFPGPSK-GDK 1185
Qy 2209 PNKTSV-----LGG-----GEDGIEPVSPGTEFGHSRVAVYPLLYRDGE---QTE 2253
Db 1186 GSKGEVGPGLAGSPGIFGVKGEQF--MGPPGQGPGLPGTPGHPVEGPKGDRGPQG 1243
Qy 2254 PSRMGSKSPGNTSPPAPFSLKTSNSAMVSKQKEIN-KKLNTHNRNEPEYNISQPGTE 2312
Db 1244 PGLFCHPGF---MGPPGF-----PGINGPKGKGNQGW---GAPGVP 1280
Qy 2313 -----IFNMPAITGTGLMTYRSQAVQEHASTNGLAII-----RKALMG---KYDOW 2357
Db 1281 GPKGDGPGGPGGIGSGPGIT-----GSKGDMGLPGVPGFQGGKGLPGLQGVKGDQ 1332
Qy 2358 EESPLSANAFNPLNASLPAAMPITAADGRSDHLLTSPGCGGKAKVSGRPSRKAASP 2417
Db 1333 DQG-----VPGPKGLQGPPGPPGPDVYIKG 1357
Qy 2418 APGLASGDRPSSVSHSEGCNRRTPLTNRV-----WEDRPSASGT-PP-PYN 2465
Db 1358 EFGLPFGPGPLGKLGLOGPPGKGGQGVTVGSLPGFPGFPGAPQKGTGTPGPG 1417
Qy 2466 PLIMRLQGVNASPPPLGAPGSLAGP-----HHAWDEPKPLIC 2507
Db 1418 P-----RGFPFGPDGLPSMGSPGTPSVDHGFLVTRHSQTTDDP---LC 1460

RESULT 58

KI67 HUMAN

ID_KI67 HUMAN STANDARD; PRT; 3256 AA.

AC P46013;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

15-MAR-2004 (Rel. 43, Last annotation update)
Antigen KI-67.
MKI67.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=94043435; PubMed=8227122;
Schulze C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
Frad H.-D., Gerdes J.;
"The cell proliferation-associated antigen of antibody Ki-67: a very
large, ubiquitous nuclear protein with numerous repeated elements,
representing a new kind of cell cycle-maintaining proteins.";
J. Cell Biol. 123:513-522(1993).
[2]
SEQUENCE OF 1-31 FROM N.A.
Gerdes J.;
Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: Thought to be required for maintaining cell
proliferation.
-!- SUBCELLULAR LOCATION: Nuclear. Predominantly localized in the G1
phase in the perinuclear region, in the later phases it is also
detected throughout the nuclear interior, being predominantly
localized in the nuclear matrix. In mitosis, it is present on all
chromosomes.
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Long;
IsoId=P46013-1; Sequence=Displayed;
Name=Short;
IsoId=P46013-2; Sequence=VSP 004298;
-!- DEVELOPMENTAL STAGE: Expression of this antigen occurs
preferentially during late G1, S, G2 and M phases of the cell
cycle, while in cells in G0 phase the antigen cannot be detected.
-!- SIMILARITY: Contains 1 FHA domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X65550; CAA4519.1; -;
EMBL; X65551; CAA4520.1; -;
EMBL; X94762; CAA64388.1; -;
PIR; A48666; A48666.
Gene; HGNC:7107; MKI67.
GK; P46013; -;
MW; 176741; -;
GO; GO:0008283; P:cell proliferation; TAS.
GO; GO:0000074; P:regulation of cell cycle; TAS.
InterPro; IPR000253; FHA.
InterPro; IPR008984; SWAD_FHA.
Pfam; PF00498; FHA; 1.
SMART; SM00240; FHA; 1.
PROSITE; PS50006; FHA_DOMAIN; 1.
Cell cycle; Antigen; Nuclear protein; ATP-binding; Repeat;
Alternative splicing; Polymorphism.
DOMAIN 27 76
DOMAIN 1000 2928 16 X 122 AA APPROXIMATE REPEATS.
FT REPEAT 1000 1112 1.
FT REPEAT 1122 1234 2.
FT REPEAT 1244 1356 3.
FT REPEAT 1366 1477 4.
FT REPEAT 1487 1598 5.
FT REPEAT 1608 1720 6.
FT REPEAT 1730 1842 7.
FT REPEAT 1851 1964 8.
FT REPEAT 1974 2086 9.

FT REPEAT 2096 2204 10.
FT REPEAT 2214 2326 11.
FT REPEAT 2235 2447 12.
FT REPEAT 2457 2569 13.
FT REPEAT 2579 2689 14.
FT REPEAT 2699 2808 15.
FT REPEAT 2818 2928 16.
FT NP_BIND 3034 3041 ATP (POTENTIAL).
FT VARSPLIC 136 495 Missing (in isoform Short).
FT VARIANT 3150 3150 /FTID=VSP 004298.
FT VARIANT 3217 3217 T -> S (in dbSNP:11106).
FT VARIANT 3217 3217 /FTID=VAR 014858.
FT VARIANT 3217 3217 K -> E (in dbSNP:8473).
SQ SEQUENCE 3256 AA; 358741 MW; 578F8C51BED42517 CRC64;
Query Match 2.6%; Score 348; DB 1; Length 3256;
Best Local Similarity 18.7%; Pred. No. 0.00032;
Matches 537; Conservative 352; Mismatches 1129; Indels 852; Gaps 133;
4 STQLVAQTWRATEPRYPHSLSYVQIARTHTDVLGLLEYQHHSRD-YAS-----51
325 SVQTPSKAVGASFLYPAKMTQVYQQNS----PQKHKNKDLTYTGRRESVNLGS 380
52 -----HLSPGSIQORRPSLL-----SEFOPGNERSOELHLRPESHSLP---EL 95
381 EGFKAGDKTLTPRK-LSTRNRTPAKVEDAADSATKPNLSKTRGSIPTDVEVLPTETEI 439
96 GKSEMEFIESKRPLELDPDLRLPSPL-----LATQPAGS-----EDLTORS 141
440 HNEPPLTLWLTOVERKIQKDSLSKPEKLTAGQMCGLPLGSSVDINNFGDSINESGI 499
142 TGKLEPVPPSPPTDPEL--ELVPPR-----LSKEELIONMDRVDREITWVEQOI 190
500 PLKRRVS--FGHLRPELFDENLPNTPLKRGAPTKEKSLVHMTFPVPLKKIIEQOPQ 557
191 SKLKKKQOOLBEE-----AKPPEKPVPPPIESKHSILVQIYDENRKKAFAAHRI 244
558 SGKQESGEIHVEVKAQSLVSPAPSPRKTVPASDQRRSCKTAPASSSKSQTEVPKR- 616
245 LEGGLGPVELPLYNQPSDTQYHENIKINQAMRKKLILYFKRNHA-----BKWKQK 297
617 ---GGERVATCLQKRVSLRSQHDILQ-----MICKRRRSASSEANLIVAKSWADV 664
298 FCORYDQLEALEKKVEIENPPRRRAKESKVREYIERQFPEIRKQRELOERMOSRVQOR 357
665 V-----KLGAKQTQTVIKHPQR-----SMNKRQRRPATPKPVGEV 702
358 GSGLSMSAARSEHEVSEIIDLSEONLEKOMRQLAVIPMLYDADQOQRIKPINNGLMA 417
703 HSQFSTGHANSP---CTIIIGKAHTEKVVHVPARPYRVLLNFI---SNQKMDFKEDLSGIA 756
418 DPMKVYKDRVM-----NMWSEQEK--ETFEKFMQHPKNFGLIASFLER--460
757 EMPKTPVKEQPOLSTCHIAISNSENLLGKQFGTDSGEELPPTSESFGNVPFSAQNA 816
461 -----KTVAECVLYYLTKNENKSLVRSYRRRGKSQOQOQOQOQOQOQOQOQOQO 514
817 AKQPSDKCSAPPLRQRCIRENGVNAKTPRNTYKMTSL-----TK 857
515 SSQBEKDEKEKEBAEKEEPEVENDKEDLLKEKTDTSGEDNDEKAVASKGRKTANS 574
858 TSDTETESKTVSTVNRSGRSTEFNRIQKLPVESKSEETNTE---IVECILKRGQKATLL 914
575 QGRKGRITRSMANESEBIAITPOOSAEALSMELNERSRTEREMETAKGLLEHGRNW 634
915 QORREGEN-----KEIRPEPTYK-ENIELKEN-----DEKMKAMKG-----SFTW 954
635 ---SAIARMYGSKTVSCKNFYNYKRONLDELTOOKLWKERNARKKKKAPAAA 690
955 GQKCAPMSDLDLKSPLPTE-LMKDTARGQNLQ-TQDHAKAPKSEKG---KITMPCQS 1009
691 SE-EAAPPVVEDEMEAS-GVSGNEEBMVBEEALHASG-----NEVPRGECSPATVN 743

Db 1010 LQPEPINTPTHTKQOLKASLGKVGKELLAVGKFRTRTSGTTHTHREPAGDGKSIRTFK 1069
QY 744 NS-----SDTESI-----PSPTAAAKD--TGONGP 767
Db 1070 ESPKILDPAAARVTGMKKWPRTPKEEAQSLEDLAGFKELFOTPGPSESMTDEKTKTKAC 1129
QY 768 KPPATLGADGPPGPPTPRRTSRAPISP-----TPASEATGAPTPPPAPPSPSPAPP 819
Db 1130 KSPPPESVDTSTKQMPKRSRLKADVBEFLALRLKLP-SAGKAMLTPKPA-----1180
QY 820 PVVKEEKEEETAAPVVE-----EGEEQKPPAAEELA-----VDTGKA 858
Db 1181 ---GGDEKDIKAFMGTPVQKLDLAGTUPGSKRQLQTPKEKAQALEDLAGFKELFOTPGHT 1237
QY 859 BEPV-----KSEC-----TEEAEEGPAKGD--AEAAEATAEGALKAEKK--EGSGKA 903
Db 1238 EELVAAGKTKIKPCDSQSDPDVDTPTSTKQPKRSIRKADVEGELLACRNLMPSAGAKWH 1297
QY 904 TTAKSSGAPOQSDSSATCSADEVDEAEG--GDKNRLLSPRPSLLTPTGDPANASPKPL 961
Db 1298 TPKEPVGEEKDIIIFVGTVPQKLDLTENLTGSKRR-----PQTPK 1337
QY 962 DLKQLKQRAAIPPIQVTKVHEPPREDAAPTKAPPAPPONLOPESDAFOQPGSSPRG 1021
Db 1338 EEAQLEDLTGFKELFOTPGHTEEAVAAGTKTKMPCSSPPES-----ADTPTS 1386
QY 1022 KSRSPAPPADKEAFAAE-----AQKLPDPPCWTSGLPFPVPPREVTKA 1065
Db 1387 TRRQPKTPLEKRDVQKELSALKKLTQTSGETHTDVKPGGDKSINAF-----RETAK- 1439
QY 1066 SPHAPDPSAFSYPAPGHP---LPLGLHDTA-----RPVLPRPPTISNPPPLISSAKH 1114
Db 1440 --QKLDPAASVTGSKRHPKTKKAQPLEDLAGWKELFOTPVCTDKPTTHEKTTKIACRSQ 1497
QY 1115 PSVLERQIGALSQMSVOLHVPYSE-----HAKAPVGPVTGVLPLP--MDPKLAPFSGV 1167
Db 1498 PDVDTTSSKPSQSKSLRKVDVEEFALRKRTPSAGKAMHTPKPAVSGEKNIYAFMG 1557
QY 1168 KQEQLS--PRGQGPPEISLGVPTAQEASVLRTGALSVPVGGSIITKIGIPSTRVPSDSAITVR 1226
Db 1558 PVQKLDLTENLTGSKRLRLOTPKE-----KAQLEDLAG-----FKELFOTRGHTEE- 1603
QY 1227 GSITHGTADLYKGTITRIIGEDSPSRDLRGREDSLPKGHVIEYEGKXHLVSEYEGMSV 1286
Db 1604 -SMINDKTAKVACKSSQDL--DKNPASSKRLKTSLGK-----GVKVELLAVG--KL 1652
QY 1287 TQCKEGRSSSGPPE-----TAAPKRTYDMHEGRVGRRAISSASIEGLMGRAPPER 1339
Db 1653 TQTSGETHTHTTEPTGDKSMKAFMESPKQLDLSAASLTG---SKRQLRTPKGSEVED 1709
QY 1340 -----HSPHLLKEQHHRGSIIT--QGI PRSYVEAQEDYLRLREAKLLKREGTPPPPPP 1389
Db 1710 LAGFIELFOTPSHYKE-----SMNKTQTKVSTRASQPD-----LVDTPTSSKKQOP 1755
QY 1390 SRDLTEAYKTOALGFLKLPAGEGLVATVKE--AGRSIHEIPIREELRHTPELPLAPRPL 1446
Db 1756 KRSLRKA-----DTEEEFLAKRQTPSAGAKWH-----TPKPA 1788
QY 1447 ---KEGSIQTGTP-LKYDTGASTTGSKKH-----DVRSLTGSPTPTPP 1486
Db 1789 VGEKDIINTFLGTVPQKLDQGNLPGSNRRLOTRKEKAQALEELTGFELEFQTPCTDNPT 1848
QY 1487 V-----HPLDVMA-----RALERACVYESLSKSPGCTASSSGGSIGARGA 1526
Db 1849 ADEKTKKILCKSPQSDPADPTPTNKORPKSLKKADVEEFLAFRKLTPSAGAKAMHTPK 1908
QY 1527 PVIVPE-----LGKPRQSPLTYEDHGAPFAGHLPRGSPVTRBPTPRLOEGLSSSSKAS 1580
Db 1909 AAVGEEKDINTFVGTPEK-----LDLLGNLP-GS--KRRPQTPK-----EKAKAL 1951
QY 1581 QD-----RKLSTP-----REIA-KSPHSTVPEHHHPHPIPSVYEHLLRGVSGVDL 1623

```
Db 1952 EDLAGKELFQTPGHTESMTDDKITEVSKSPQ---PDVVKTPPTSQKRLKISLGKGV 2008
QY 1624 YRSHIPLA-FDPTSIPRGIPLDAAAAYLPRHLAPNPTYPHYLYRGYPDTAALNR 1682
Db 2009 KEVLVPVGLTOTS-----GKTTQTH-----R 2030
QY 1683 QTIINDYITSQOHHNTATAMQADMLRGL-----SPRESSALNAAAGPRGIIDLSQV 1737
Db 2031 ETAGDG--KSIKAFKESAKQMLDPANYGTGMERWPTPKKEAQSLEDLAGFK--ELFQT 2085
QY 1738 P-HLPVLPVPTGCTATANDRLAYLPTAPQPPSSRHSSPLSPGPGTHLTPTTSSSER 1796
Db 2086 PDHT-----ESTTDDKTTKJACKSP-----PEMSDPTST-----R 2118
QY 1797 ERDRERDRDREREKSI---LTSITTTVEHAP-----IWRPCTEOSSGSSGSGGG 1845
Db 2119 RPKPTPLGKRDIVEELSALKQITQTHDKVPGDEDKGINVFRETAKQLDPAASVTG-- 2176
QY 1846 GSSRPASHSHAHQHSPIPRPTQDALQORPSVLHNTGMKGII-TAVEPSKPTVLSTSTS 1904
Db 2177 -----SKQRPRTPKGAQPLEDL--AGLKELFQTPVCTDKPTTHEKT-TX 2218
QY 1905 SPVRPAANTFPATHCPLGTLTGCVPTLMEP-----VLLPKAPRVARP- 1948
Db 2219 IACRSFPQDPVGT-----PTTFKPSQSKSLRKADVEESLALRKRTSPVSKAM 2266
QY 1949 ERPRADTG-----HAFKAPPARSGLEPASPSKSGSEPRPLVP-----PVSGHATIA 1995
Db 2267 DTPKAGGDKMKAFTMGTPVOKLID-PGNLP--GSKWPQTPKEKAQLEDLAGFKELF 2323
QY 1996 RTPAKN-----LAPHASPPPPAPPASADPHREKTSQSPFSIQE--LELRLSGY 2043
Db 2324 QTPGDKPTTDEKTKTKIACKSPQDPVDTPASTK--QRPKRLRKADVEEFLALR--- 2377
QY 2044 HGSSVSPGVEPVSVPSPSLTHDKGLPKHLELDKSHLEGELRPKQCPVKLGEEAHL 2103
Db 2378 ---KRTPSAGKAWD--TPPAVSDEKNIINTFVET-----PVOKLIDLGLNL 2417
QY 2104 PHRLPSPQSSPSSPLLOTAPGVK-----GHQRVVTLAQHISEVITQD-----YTRH 2150
Db 2418 PGSKRQPTPKREKAEALDVLGFKELFQTPGHTESMTDDKITEVSKSPQSPESFKTSR 2477
QY 2151 HPQLSAFLPAPLYSPFGASCPLDLRRPP---SDLYLPPPD---HGAPARGSPHSEGG 2203
Db 2478 SKQRLKIPLVK-----VDMKEPLAVSKLRTSGTETQTHTEPTGDSKSIKAF 2525
QY 2204 KRSPB---PNKTSVLGGGED-----GIEP-VSPPEGMTEPCHSRSAVYPLLYRDE 2250
Db 2526 KESPAQILDPAASVTGSRQRLTRKEKARALEDLVDFKELFSAPGHTESM-----T 2577
QY 2251 QTEPSRMGSKPGNTSQPPAFPSKLTESNAVKSKOEINKKLNTHRNPEYNI SQPG 2310
Db 2578 IDKNTKICKSP-----FPFLTDATSTKRCPKTRPRKEVKEELSAVERLT---QTSGQS 2629
QY 2311 TEIFNMPAITGTLMTYSQA-----VOEHAS----- 2337
Db 2630 THTHKEPASGDEGKVLKQRAKKNPVEEPSPRRRRPRAPKEKAQPLEDLAGFTLSETS 2689
QY 2338 --TNMGLEAIRKALMGKYDOWE--ESPPLSANAFNPLNASAL-----P 2378
Db 2690 GHTQESLTA-----GKATKIPCESPLEV-----VDTTASTYRHLTRVQKVQVKEEP 2737
QY 2379 AAMPITAADGRSDHTLTPSGGGGKAKVSGRPSRRKAKSPAPGLASGRPP 2428
Db 2738 SAVKFTQTSGETTODAKEPAGEDKGIKALKESAKQTPAPAAASVTGSRRRP 2787

RESULT 59
ID ACIN MOUSE STANDARD; PRT; 1338 AA.
AC Q9JIX8; Q9CSN7; Q9CSR9; Q9CSX7; Q9R046; Q9R047;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
```

```
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Apoptotic chromatin condensation inducer in the nucleus (Acinus).
GN ACINUS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RX MEDLINE=99418556; PubMed=10490026;
RA Sahara S., Aoto M., Eguchi Y., Inamoto N., Yoneda Y., Tsujimoto Y.;
RT "Acinus is a caspase-3-activated protein required for apoptotic
RT chromatin condensation.";
RL Nature 401:168-173(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX Mamoru A., Setsuko S., Yoshihide T.;
RT "Molecular cloning of murine acinusl, a gene for apoptotic chromatin
RT condensation.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-1190 FROM N.A. (ISOFORMS 1; 2 AND 4).
RC STRAIN=CS7BL/6J; TISSUE=Embryo, and Pancreas;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.B., Cousins S.,
RA Dalla E., Dragani L., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Gimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempie C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilmung L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
CC -!- FUNCTION: Induces apoptotic chromatin condensation after
CC activation by CASP3 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=1;
CC IsoId=Q9JIX8-1; Sequence=Displayed;
CC Name=2; Synonyms=5;
CC IsoId=Q9JIX8-2; Sequence=VSP_004030, VSP_004033;
CC Name=3; Synonyms=5;
CC IsoId=Q9JIX8-3; Sequence=VSP_004031;
CC Name=4;
CC IsoId=Q9JIX8-4; Sequence=VSP_004032;
CC -!- PTM: Undergoes proteolytic cleavage; the processed form is active,
CC contrary to the uncleaved form (By similarity).
CC -!- SIMILARITY: Contains 1 SAP domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC in position 110 and 112.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



```
Db 987 GVSITIDDPVTAQVSPSPRGKISNIVHISNLVRPFTLLQKELLGRTGTLVEAFWIDK 1046
QY 1459 -----YDT--GASTTCKSKHVDYSLIGSPRTFPVPHPLDVWADARALACYEESL 1508
Db 1047 IKSHCFVTYSTVEEAVATKTLHGK-----WQSNPKFLCADYAEQDELHYH--- 1094
QY 1509 KSRPGTASSGGSGIARGAPVIVPELKGKQPSPLTYEDHGAPGAPHLGRGSPVTWREPTP- 1567
Db 1095 -----RGLLV-----DRPSETKAEQAGAPRPLHPPPPPV---QPPPH 1129
QY 1568 -----RLQSGSLSSKASQDR-----KLSTPR-----ETAK 1594
Db 1130 PRAEQEGERAVREQWAEREREMERRTSRERWDKVRGPRSRGRSDRRRKRERAK 1189
QY 1595 SPHSTVPEHHPPISPYPHELLRGVGVLYRSHIPLADPTSPICGIPLDAAAAYLPRH 1654
Db 1190 SKEKSEKKEKAQEPKAKLLD-----DLFRK-----TKAAPCIYWLp-- 1227
QY 1655 LAPNPTYPHYLPYLLIRGYPDTAALENQTIINDYITSQQMHNTATAMQADMLRGLS 1714
Db 1228 -----LTSQIVQKEAQERAKEREKRRK 1252
QY 1715 PRESSLALNAAAGPRGIIDLQVPHLPVLVPTPGTATAMDRLAYLPTAPQPFSSRHSS 1774
Db 1253 EREEE-----EQKER-----EKEAERERN 1271
QY 1775 SPLSPGGTHLTKPTTSSSRERDRDRDR-DREREK 1812
Db 1272 RQLE-----REKRSRERDRERDRERDRER 1303

RESULT 60
APC_MOUSE
ID APC_MOUSE STANDARD; PRT; 2845 AA.
AC Q61315; Q62044;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein) (MAPC).
GN APC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.
RC STRAIN=C57BL/6J, and CAST/EI; TISSUE=Brain;
RX MEDLINE=92263101; PubMed=1350108;
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RT "Multiple intestinal neoplasia caused by a mutation in the murine
RT homolog of the APC gene.";
RL Science 256:668-670(1992).
RN [2]
RP ERRATUM.
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RL Science 256:1114-1114(1992).
RN [3]
RP SEQUENCE OF 1-45 FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Dicker F., Lambertz S., Reitmaier A., Ballhausen W.G.;
RT "The murine APC gene: alternative splicing of 5' untranslated
RT region segments.";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
ALTERNATIVE SPLICING.
RX MEDLINE=94061824; PubMed=8242607;
RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
RA "APC gene messenger RNA: novel isoforms that lack exon 7.";
RL Cancer Res. 53:5589-5591(1993).
CC -!- FUNCTION: tumor suppressor. Promotes rapid degradation of CTNNB1
and participates in Wnt signaling. APC activity is correlated with
```

```
CC CC its phosphorylation state (By similarity).
CC CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
CC CC axin (By similarity).
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC Event=Alternative splicing; Named isoforms=4;
CC CC Name=1;
CC CC IsoId=Q61315-1; Sequence=Displayed;
CC CC Name=2;
CC CC IsoId=Q61315-2; Sequence=VSP_004116;
CC CC Name=3;
CC CC IsoId=Q61315-3; Sequence=VSP_004117;
CC CC Name=4;
CC CC IsoId=Q61315-4; Sequence=VSP_004116, VSP_004117;
CC CC -!- TISSUE SPECIFICITY: Expressed in liver, spleen, kidney, heart,
CC CC lung, brain, stomach, intestine, testis and ovary.
CC CC -!- PTM: Phosphorylated by GSK3B (By similarity).
CC CC -!- SIMILARITY: Contains 7 ARM repeats.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@sib-sib.ch).
CC CC
CC CC EMBL; M88127; AAB59632.1; -.
CC CC EMBL; U02937; AAA03443.1; -.
CC CC PIR; I49505; I49505.
CC CC HSP; Q02248; 3BCT.
CC CC
CC CC MGD; MGI:880339; Apc.
CC CC GO; GO:0005737; Cytoplasm; IDA.
CC CC GO; GO:0005634; C:nucleus; IDA.
CC CC GO; GO:0008013; F:beta-catenin binding; IDA.
CC CC GO; GO:0009952; P:anterior/posterior pattern formation; IMP.
CC CC GO; GO:0009798; P:axis specification; IMP.
CC CC GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.
CC CC GO; GO:0016055; P:Wnt receptor signaling pathway; IDA.
CC CC InterPro; IPR008938; ARM.
CC CC InterPro; IPR000225; Armadillo.
CC CC Pfam; PF00514; Armadillo_seg; 4.
CC CC SMART; SM00185; ARM; 5.
CC CC PROSITE; PS0176; ARM_REPEAT; 1.
CC CC Wnt signaling pathway; Anti-oncogene; Phosphorylation;
CC CC Alternative splicing; Repeat; Coiled coil;
CC CC DOMAIN 1 61 COILED COIL (POTENTIAL).
CC CC DOMAIN 125 245 COILED COIL (POTENTIAL).
CC CC REPEAT 1 728 LEU-RICH.
CC CC REPEAT 451 493 ARM 1.
CC CC REPEAT 503 545 ARM 2.
CC CC REPEAT 546 589 ARM 3.
CC CC REPEAT 590 636 ARM 4.
CC CC REPEAT 637 681 ARM 5.
CC CC REPEAT 682 723 ARM 6.
CC CC REPEAT 724 765 ARM 7.
CC CC DOMAIN 739 2834 SER-RICH.
CC CC DOMAIN 1130 1156 ASP/GLU-RICH (ACIDIC).
CC CC DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
CC CC DOMAIN 1864 1891 HIGHLY CHARGED.
CC CC VARSPLIC 243 276 Missing (in isoform 2 and isoform 4).
CC CC VARSPLIC 310 410 Missing (in isoform 3 and isoform 4).
CC CC VARSPLIC 120 410 /FTId=VSP_004117.
CC CC VARSPLIC 120 410 /FTId=VSP_004117.
CC CC VARIANT 120 120 T -> A (IN STRAIN CAST/EI).
CC CC VARIANT 493 493 Y -> I (IN STRAIN CAST/EI).
CC CC VARIANT 797 797 Y -> F (IN STRAIN CAST/EI).
CC CC VARIANT 1330 1330 A -> T (IN STRAIN CAST/EI).
CC CC VARIANT 1618 1618 A -> S (IN STRAIN CAST/EI).
CC CC VARIANT 2294 2294 G -> A (IN STRAIN CAST/EI).
CC CC VARIANT 2496 2496 H -> Q (IN STRAIN CAST/EI).
CC CC VARIANT 2523 2523 T -> A (IN STRAIN CAST/EI).
CC CC VARIANT 2813 2813 T -> S (IN STRAIN CAST/EI).
CC CC SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A99 CRC64;
```



```
Db 243 GOIQSGAAQMAQETRDALKCDVTSALRE---IRAQLEGHAVOSTLQSEWFRVRLDR 299
Qy 305 LMEALEKKVRIENPRRAKESKVEYYEQFPEIRKQRELQRMOSRVQQRGSLMS 364
Db 300 LSEAAK-----VNTDAMSQAETIEY-----RRQLQARTTELEALKSTKOSLE 343
Qy 365 AARSEHVESEIIDGLSEENLEKOMROLAVIPMLYDADQOQRIKFINNGIAMDPMKYK 424
Db 344 RQSELEDHRHODIASYQEAQQ-----LDAELRNTKW-----EMAAQLREYQ 386
Qy 425 DRQVMNMWSEOEKETFREKFWHPK---NFGLIASFLEKTVAECLVLYLYLTKNENYKS 481
Db 387 DLLVMKQALDIEIAAYR-KLLEGEBCRIGFGPIFSLPEGLPKIPSVSTHIKXSEKIK 445
Qy 482 LVRSYRRRGKSQQQQQQQQQQQQQQPPRSQEEKDEKEKEKEKEKEKEKEPEVEND 541
Db 446 VVESEKETVIVEQTEBTQTEEV-----TEEEKEKEKEKEKEKEKEEBAEGG 497
Qy 542 KEDLLKEKTDGTDGNDKEKAVASKGRKKTANSQGRKGRITRSMANEANSEEAITP--- 598
Db 498 BEETKSPPAEBAASPEKAKSPVKEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAKE 556
Qy 599 -----QSSAELASNELNNESSRWTE-----EMETAKKGLLEHGRNWS 635
Db 557 EAKSPPEAKSPKEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAE 616
Qy 636 AIARWVGSKTVSQCNFYFNFKKQNLDEILQOHKLW-----EKERNARRKKKAPA 688
Db 617 EAKSPVKEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPV 676
Qy 689 ---AASEEAPPPVVEDEMBASVSGNEEEMVEAEALHASGNEVPRGECSPATVNN 745
Db 677 KAEAKSPAEKSPVKAEBKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPVKE 736
Qy 746 SDTESIPSPHTEAAKDTQ--NGPRPPTLGDGPPGPPPTPRTSRAPTEPTASEATG 804
Db 737 KTPEKAKSPVKEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAE 792
Qy 805 APTPPAPPSPAPPPVVPKEKEEETAAAPVVEGEGBQKPPAAELAVDTGKAEPEVK 864
Db 793 SPVKEEVKSPKAKSPL-----KEDAKAPEKEIKPEKEVKSPVKEEKPOEVKEPPKK 847
Qy 865 ECTEAEGBPAKDAEAAETAGALKAEKKEGSGRATTAKSGAPQDSDSATCSAD 924
Db 848 ---AEEKAPATPKTEKKDSKE---EAPKKEAPKPKVEKSGPAVEKPKSKV----- 896
Qy 925 EVDEAEGDKRLLSPRSLTPTGDPDRANASPOKPLDLKQLKORAAIPIQVTKVHEP 984
Db 897 EAKKEAEDKKV-----PTPEKAPAKVEKEDAK-----PKETEVAKK 937
Qy 985 PRED---APTKAPPAPPPQNLOQESDAPQPGSSPRGK-----SRSPA-PPA 1030
Db 938 EPDDAKAKEPKPAKKEKAEAEKDKTKEKAKKPEKPKTEAKAKEDKTLKSPKPKA 997
Qy 1031 DKEAFPAEAKLPGDPP 1047
Db 998 EAKESSTQDKSKPP 1014

RESULT 62
CALL MOUSE
ID CALL MOUSE STANDARD; PRT; 1453 AA.
AC P11087; Q60635;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1 OR COLA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

```
RP RC SEQUENCE FROM N.A.
RX STRAIN=FVB/N;
RX MEDLINE=96033240; PubMed=8535610;
RA Li S.W., Khillan J., Prockop D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
of type I procollagen.";
RL Matrix Biol. 14:593-595 (1995).
RN [2]
RP SEQUENCE OF 518-1128 FROM N.A.
RX MEDLINE=86137403; PubMed=3841523;
RA French B.T., Lee W.-H., Maul G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
collagen protein.";
RL Gene 39:311-312 (1985).
RN [3]
RP SEQUENCE OF 735-1130 FROM N.A.
RX MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
evidence for a mouse B1 element within the gene.";
RL Mol. Cell. Biol. 2:1362-1371 (1982).
RN [4]
RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX MEDLINE=83157109; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences.";
RL DNA 1:59-69 (1981).
RN [5]
RP SEQUENCE OF 1442-1453 FROM N.A.
RX MEDLINE=88124276; PubMed=3340560;
RA Moolenaar K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
of the 3'-untranslated region.";
RL Nucleic Acids Res. 16:773-773 (1988).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
(fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
bones. In bones the fibrils are mineralized with calcium
hydroxyapatite.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- SIMILARITY: Contains 1 VWF domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U08020; AAA88912.1; -
DR EMBL; X15896; CAA33904.1; -
DR EMBL; M14423; AAA37333.1; -
DR EMBL; M17491; AAA37334.1; -
DR EMBL; X06753; CAA29927.1; -
DR EMBL; K03036; AAA37332.1; -
DR EMBL; K03029; AAA37332.1; JOINED.
DR EMBL; K03030; AAA37332.1; JOINED.
DR EMBL; K03031; AAA37332.1; JOINED.
DR EMBL; K03032; AAA37332.1; JOINED.
DR EMBL; K03033; AAA37332.1; JOINED.
DR EMBL; K03034; AAA37332.1; JOINED.
DR EMBL; K03035; AAA37332.1; JOINED.
DR PIR; S57243; S21626.
DR MGI; MGI:88467; Col1a1.
DR InterPro; IPR008161; Ctg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib-collagen_C.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR001007; VWF_C.
```

DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; C1g helix; 1.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VMC; 1.
 DR PROSITE; PS01208; WFEC 1; 1.
 DR PROSITE; PS0184; WFEC 2; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 151
 FT CHAIN 152 1207
 FT PROPEP 1208 1453
 FT DOMAIN 29 87
 FT DOMAIN 152 167
 FT DOMAIN 168 1181
 FT DOMAIN 1182 1207
 FT CARBOHYD 56 56
 FT CARBOHYD 1354 1354
 FT SITE 734 736
 FT SITE 1082 1084
 FT CONFLICT 1450 1450 A -> V (IN REF. 5).
 SQ SEQUENCE 1453 AA; 137944 MW; 3802E535DF81808 CRC64;

Query Match
 Best Local Similarity 21.7%; Pred. No. 0.00025;
 Matches 335; Conservative 89; Mismatches 539; Indels 580; Gaps 80;

QY 707 ASGVSGNBEEMVEAEALHASQNEVPRGE-----C-SGDAVNNSSDTEISIPSP 754
 DB 15 ATALLTHQOEIDPEVSCIH-NGLRVPNGETWKPEVCLICICNGHTAVCDVQCNEELDCP 73
 QY 755 HTEA-----AKDTGQNGPK-----PPATLGADGPP--PGP 782
 DB 74 NQRRREGCCACPREYVSPNSDVGVEGKGGPQGPQGPVGGPRDGIIPGQGLGPP 133
 QY 783 PTPPRRTSRAPIETPASEATGATPP-----PAPSPSAP- 818
 DB 134 PGPP-----GPPPGPLGGNFASQMSYGVDEKSAQSVFPGMGPSPGR 176
 QY 819 ---PPVVPKKEBETAAPVEGE-----EQKPPAAELAVDTGKAEPVVKSEC 866
 DB 177 GLGPPGAPGQFGQ-----GPPGEPGPGGMPGPPGPPGKNGDGDGAGKPR- 229
 QY 867 TEEAEGPAKGDAAEATAEGALKAEKKGSGRATTAKSSGAPOQSDSSATCSADEV 926
 DB 230 ---PGERGPPQAGRLPCTA--GLPGMKHGRGSLDGAKGDAGP----- 271
 QY 927 DEAEQGDKNRLSP-----RPSLLTPTGDPANASPKQLDLKQLKQRAAAIPPIQVTKVH 982
 DB 272 ---AGPKGEPGSPGNGAPGQMGPRGLPGERGRGPP-----GTAGA-----RGN 313
 QY 983 EPPREDAAPTKAPAPP-----PQNLPESDAPO-----QPSSPRGKRS 1025
 DB 314 DAGVAAGPGPTGTGTPGPFVGAAGKAGAPQCAR--GSEGPQGVGEPGPPGAPGAAG 372
 QY 1026 PA--PPADKEAFAEAQKLPDPPCWTSGLP-FVPPPREVIKASPHAPDPSAFYAPPGH 1082
 DB 373 PAGNPGADGQPGAKGANGAPG-----IAGAPGFCARGPSPQCPGPPGKNGSGEPGA 427
 QY 1083 PLPLGLHDTARPVLPPTTINPP-PLISSAKH-----PSVL-----ERQIGAISQGM 1129
 DB 428 PGNKG--DITGAKGPFATGVQGPAPAGEEKGKRGARGEPGPGSLPGPPGERG--GPGSRGF 484
 QY 1130 SVQLHVPYSEHAKAPVGVV-TWGLPLPMDPKLAPFSVQEQLSPRGQAGPPESLVPT 1188
 DB 485 -----FCADGVAGPKGSGERGAPGAPGPKG-----SP-GEAGRPGEGALPG 525
 QY 1189 AQEASVLRGT-----ALGSVPGGSIKGIPTSPVPSDAITVRGSI--THGTADVLYKGTI 1243
 DB 526 AKGLTSGSPGPPDGKTPGPPGAPQDGRGPPGAPPGA-----RGQAGVMGFP--KGT- 577

RESULT 63

SHK3 RAT
 ID_SHK3_RAT STANDARD; PRT; 1815 AA.
 AC OSUJIA; Q9WU7; Q9WU47;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE SH3 and multiple ankyrin repeat domains 3 (Shank3) (Proline-rich
 DE synapse associated protein 2) (ProSAP2) (SPANK-2).
 GN SHANK3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

QY 1244 TRIIGEDSPSL-DRGREDSLKPHVIEGKGKGVLSVEGMSVTCQSKEDGRSSSGPPH 1302
 DB 578 ---AGE--PGKAGERGLPG--PPGAVGAPAGDG-----EAGAQAAGPAGPAG 618
 QY 1303 ETAAPKRTYDMMEGRVGRAISSASIEGLMGRAIIPP-ERHSPHILKEQHHRGSIQTGIPR 1361
 DB 619 E-----RGEQGA--GSPGFQGLPGPAGPPGEAKPGE-----QGVPG 654
 QY 1362 SYVEAQEDYLREAKLLKREGTPPPPPPSRDLTEAYKTQALGPL--KLKPAHEGLVATVK 1419
 DB 655 DLGAPGPGSGARGERGPFCEGVQVQPPGA-----GPRNGAPGNDCAKGDG 702
 QY 1420 EAGRSIHEIPRELRHTPELPLAPRLKEGSIQTGP-----LKDYDTGASITGSKHP- 1472
 DB 703 APG-----AP-----GS--QGAFLQGMPPGERGAAGLPGPKGRGD 736
 QY 1473 --VRSLLGSPGR-----TFP--PVHPLDMADARALERACYESLSRSPGTASSGSI 1522
 DB 737 AGPKGADGSPKCGARGLTGPIGPPGAPAGD-----KGEAGPSGPPGPTG 783
 QY 1523 ARGAPVIVPELGKPRQSPFLTYEDHGAPFAGHLPRGSPVTWREPTPRLOBSLSSSKASQD 1582
 DB 784 ARGAPGDRGEAGPP--GPAGFA--GPPGAD-----GQFKAKEGPDGTGVKGD- 826
 QY 1583 RKLTSPTREIAKSPHSTVPEHHPHPI---SPVEHLLRGVGVLDLYRSHIPLAFDPTSIP 1638
 DB 827 -----AGPPGAPAGPAGPGPIGNVGAAPGPKGRGAAP-----P 860
 QY 1639 RGIPLDAAAAYILPRHLPNPTYPHYLYPIRIGYPTAALENRQTIINDYITSQQMHN 1698
 DB 861 GATGFPGAAGRVG---PGFS--GNAGPP---GPPGPVKEG----- 894
 QY 1699 TATAMAQADMLRGLSPRESSLALNVAAGPRGIIIDISOVPHLPVLVPPTGTATAMDRL 1758
 DB 895 -----GKGRGET-----GPAG-----RPEVGP--FGPPFGAGEKG 924
 QY 1759 AYLTPAPQPPSSRHSSSPLSPG-----GPTHLTKPTTS 1792
 DB 925 S--PGADGP-----AGSPGTPGQIAGQGVVLGQRCGERGFPGLPGSGEPGKQGPS 977
 QY 1793 SSERDRDRDRDREREKSIITSTTVEHAIWRPTEQSSGSSGSSG-----GGGS 1847
 DB 978 GSSGREG------PPGMPGPPGLAGPPGSGREGSPGAEAGSPGR 1015
 QY 1848 SSRPASHSHAHQSPISPTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSSPV 1907
 DB 1016 DGAPGAKGRGETGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 1067
 QY 1908 RPAATFPPTHCPGLGGLDGVYPTLMEPVLLPKEAPRVARPERPRADTGH----- 1957
 DB 1068 GPAGARGP-----AGPQGRGDKGETGQDGRGIK 1097
 QY 1958 -----AFLAKPPARSL-----EPASSPKGSE-----PRPLVPPVSG 1990
 DB 1098 GHRGFSGLQPPGSPGSGQSGASGAPGPRGPPGSGAGSGDKGLGLGPGIPGPPGR 1157
 QY 1991 HATIAITPAKVLAPHASDDPPAPPASD-----PHREKTO 2027
 DB 1158 GRTGDSGPAGPPCP-PGPPGPPGPPSGGYDFSLPQPPQEKSQ 1199

[illegible][illegible]

RESULT 64
 ID _CAL3 HUMAN STANDARD; PRT; 1466 AA.
 AC P02451; Q15112;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(III) chain precursor.
 GN COL3A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RN SEQUENCE FROM N.A.
 RC TISSUE=Skin fibroblast;
 RX MEDLINE=89350838; PubMed=2764886;
 RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
 RA Prockop D.J.;
 RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
 RT chain of human type III procollagen. Differences in protein structure
 RT from type I procollagen and conservation of codon preferences.";
 RL Biochem. J. 260:509-516(1989).
 RN (2)
 RN SEQUENCE OF 149-1225 FROM N.A.
 RX MEDLINE=89386015; PubMed=2780304;
 RA Janeczko R.A., Ramirez F.;
 RT "Nucleotide and amino acid sequences of the entire human alpha 1
 RT (III) collagen.";
 RL Nucleic Acids Res. 17:6742-6742(1989).
 RN (3)
 RN SEQUENCE OF 168-398.
 RX MEDLINE=77134724; PubMed=557335;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of cyanogen
 RT bromide peptides from the amino-terminal segment of type III collagen
 RT of human liver.";
 RL Biochemistry 16:1158-1164(1977).
 RN (4)
 RN SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 RT procollagen gene (COL3A1).";
 RL Gene 78:255-265(1989).
 RN (5)
 RN REVIEW ON VARIANTS.
 RA Seyer J.M.;
 RL Submitted (DEC-1977) to the PIR data bank.
 RN (6)
 RN SEQUENCE OF 399-727.
 RX MEDLINE=79000343; PubMed=687591;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of five
 RT consecutive CNBr peptides from type III collagen of human liver.";
 RL Biochemistry 17:3404-3411(1978).
 RN (7)
 RN SEQUENCE OF 728-964.
 RX MEDLINE=80198282; PubMed=6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha 1
 RT (III)-CB5 from type III collagen of human liver.";
 RL Biochemistry 19:1583-1589(1980).
 RN (8)
 RN SEQUENCE OF 950-1466 FROM N.A.
 RX MEDLINE=88189827; PubMed=3357782;
 RA Mankoo B.S., Daigleish R.;
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
 RL Nucleic Acids Res. 16:2337-2337(1988).
 RN (9)
 RN REVISION TO 1184.
 RX MEDLINE=89098346; PubMed=3211760;
 RA Molyneux K., Daigleish R.;
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
 RL Nucleic Acids Res. 16:11833-11833(1988).
 RN (10)
 RN SEQUENCE OF 1065-1466 FROM N.A.
 RX MEDLINE=85087944; PubMed=6096827;
 RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
 RA Rosenbloom J., Myers J.C.;

RT "Molecular cloning and carboxyl-propeptide analysis of human type III
 RT procollagen.";
 RL Nucleic Acids Res. 12:9383-9394(1984).
 RN (10)
 RN SEQUENCE OF 965-1200.
 RX MEDLINE=81208139; PubMed=7016180;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha
 RT 1(III)-CB9 from type III collagen of human liver.";
 RL Biochemistry 20:2621-2627(1981).
 RN (11)
 RN SEQUENCE OF 1176-1466 FROM N.A.
 RX MEDLINE=85157600; PubMed=2579949;
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sibbala M., Ramirez F.;
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
 RT (III) collagen. Partial characterization of the 3' end region of the
 RT gene.";
 RL J. Biol. Chem. 260:4357-4363(1985).
 RN (12)
 RN SEQUENCE OF 1161-1200 FROM N.A.
 RX MEDLINE=86187804; PubMed=3754462;
 RA Miskulin M., Daigleish R., Klueve-Beckerman B., Rennard S.I.,
 RA Tolstoshev P., Brantly M., Crystal R.G.;
 RT "Human type III collagen gene expression is coordinately modulated
 RT with the type I collagen genes during fibroblast growth.";
 RL Biochemistry 25:1408-1413(1986).
 RN (13)
 RN SEQUENCE OF 1-170 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88303360; PubMed=3405773;
 RA Toman D., Ricca G., de Crombrughe B.;
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
 RT of human prepro alpha 1(III) collagen.";
 RL Nucleic Acids Res. 16:7201-7201(1988).
 RN (14)
 RN SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 RT procollagen gene (COL3A1).";
 RL Gene 78:255-265(1989).
 RN (15)
 RN REVIEW ON VARIANTS.
 RA MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN (16)
 RN VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
 RX MEDLINE=93293988; PubMed=8514866;
 RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinhert C.,
 RA Earley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,
 RA Cole C.W., Jaakkola P., Rymnen M., Pearce W.H., Yao J.S.T.,
 RA Majamaa K., Smullens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
 RA Jackson C.E., Michels V., Kaye M., Kuivaniemi H.;
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
 RT in the triple-helical domain of type III procollagen are an
 RT infrequent cause of aortic aneurysms.";
 RL J. Clin. Invest. 91:2539-2545(1993).
 RN (17)
 RN VARIANT THR-698.
 RX MEDLINE=91045136; PubMed=2235526;
 RA Zafarullah K., Kleinhert C., Tromp G., Kuivaniemi H., Kontusaari S.,
 RA Wu Y., Ganguly A., Prockop D.J.;
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
 RL Nucleic Acids Res. 18:6180-6180(1990).
 RN (18)
 RN VARIANT AORTIC ANEURYSM ARG-786.
 RX MEDLINE=91056145; PubMed=2243125;
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family

Db 916 ----SPGVSGKGDAGQCEKSPGAQGPAGPLG-----IAGI----- 952
QY 1946 ARPERPRADTHAFIAPKPARSLGPASSPS-----KGSEPRPLVPPVSGHATARTPAKN 2001
Db 953 -----TGARGLAGPPGPG--PRGSPGQGVKGSGKP-----GANG 987
QY 2002 LAPHASDPP---PAPPASADPHREKTQSKPFSIQELESRLGYHSGSSYSPGV-----E 2054
Db 988 LSGERGPPGPGQGLPLCLAGTAGCPGDRGNP-----SDGLPGRDGSFGGKGRGE 1036
QY 2055 PVSFVSSSLTHDKLPHLSELDKSHLEGLRKPQCPVKLGGAHLPHLRPLPESQP 2114
Db 1037 NGSP-GAFGAPCHPDPGPGVPGAGKSGDRGESGP--AGFAGAPGAGS-----RGAPGPGQ 1089
QY 2115 SSSPLLQT---APGVKHQRVWTLAQHISEVITQDYTRHHPPQQLSAP-LPAP-----LY 2164
Db 1090 PRGDKGTGERGAAGIKGH-----RGFPNGPAGSPGPGAGQQGAI 1130
QY 2165 SFPGASCVLRLRP-----PSDLYLPPPHDHCAPARGSPHSEGGKRSPEP 2209
Db 1131 GSPGAPG---RGVPGSPGPKDGTSGHPGPIGPPGPRGNRGSGSGSPGHPGQPGP 1186
QY 2210 -----NKTSLVLG-GGED-----GIEP-----VSPPEGMTPEGHSRSAYV 2242
Db 1187 PGPGAPGCGGCGVGAAGIAGTGGKAGGFAPYVGDEPMDPKINTDEMTSLKSVNGQIE 1246
QY 2243 PLLYRDGQTPSR 2256
Db 1247 SLSPDGSRKKNPAR 1260

RESULT 65

ACIN HUMAN
ID ACIN HUMAN STANDARD: PRT: 1341 AA.
AC Q9UKV3; OY5158; Q9UG91; Q9UKV1; Q9UKV2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apoptotic chromatin condensation inducer in the nucleus (Acinus).
GN ACINUS OR KIAA0670.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), PARTIAL SEQUENCE, FUNCTION,
RP AND MUTAGENESIS OF ASP-1093.
RX MEDLINE=99418558; PubMed=10490026;
RA Sahara S., Aoto M., Eguchi Y., Imamoto N., Yoneda Y., Tsujimoto Y.;
RT "Acinus is a caspase-3-activated protein required for apoptotic
RT chromatin condensation";
RL Nature 401:168-173(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Uterus;
RA Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Fetal brain;
RA Li W.B., Gruber C., Jessee J., Polayes D.;
RT "Full-length cDNA libraries and normalization";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 56-1341 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kokani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";

RL DNA Res. 5:169-176(1998).
CC -!- FUNCTION: Induces apoptotic chromatin condensation after
CC activation by CASP3.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=L;
CC IsoId=Q9UKV3-1; Sequence=Displayed;
CC Name=2; Synonyms=S';
CC IsoId=Q9UKV3-2; Sequence=VSP_004025, VSP_004028;
CC Name=3; Synonyms=S;
CC IsoId=Q9UKV3-3; Sequence=VSP_004026, VSP_004029;
CC Name=4;
CC IsoId=Q9UKV3-4; Sequence=VSP_004027;
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- PTM: Undergoes proteolytic cleavage; the processed form is active,
CC contrary to the uncleaved form.
CC -!- SIMILARITY: Contains 1 SAP domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF124726; AAD56724.1; -
CC EMBL; AF124727; AAD56725.1; -
CC EMBL; AF124728; AAD56726.1; -
CC EMBL; AL050382; CAB43681.1; -
CC EMBL; BX247975; CAD62309.1; -
CC EMBL; AB014570; BAA31645.2; -
CC Genew: HGNC:17066; ACINUS.
CC MIM: 604562; -
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0016887; F:ATPase activity; NAS.
CC GO; GO:0019899; F:enzyme binding; NAS.
CC GO; GO:0003676; F:nucleic acid binding; NAS.
CC GO; GO:0030263; P:apoptotic chromosome condensation; IDA.
CC GO; GO:0030218; P:erythrocyte differentiation; IEP.
CC GO; GO:0045657; P:positive regulation of monocyte differentiation. .; IEP.
CC InterPro; IPR003034; SAP.
CC Pfam; PF02037; SAP; 1.
CC SMART; SM00513; SAP; 1.
CC PROSITE; PS00800; SAP; 1.
KW Apoptosis; Nuclear protein; Alternative splicing.
FT DOMAIN 72 106
FT GLU-RICH.
FT SER-RICH.
FT PRO-RICH.
FT ARG/ASP/GLU/LYS-RICH.
FT CLEAVAGE (BY CASPASE-3).
FT Missing (in isoform 2).
FT /FTid=VSP_004025.
FT Missing (in isoform 3).
FT /FTid=VSP_004026.
FT Missing (in isoform 4).
FT /FTid=VSP_004027.
FT VARSPLIC 1 758
FT VARSPLIC 1 1152
FT VARSPLIC 728 766
FT VARSPLIC 759 766
FT MUTAGEN 1093 1093
FT CONFLICT 139 139
FT SEQUENCE 1341 AA; 151887 MW; 8FE286681F83AB5C CRC64;
GSPKKEAEAPPAATQPTSTQTSHTLPSERIHTTV
P (in isoform 2).
SERIHTTV -> MLSEKEG (in isoform 3).
D->A: ABOLISHES CLEAVAGE BY CASP3 AND
CHROMATIN CONDENSATION ACTIVITY.
Q -> H (IN REF. 4).
Query Match 2.6%; Score 339.5; DB 1; Length 1341;
Best Local Similarity 18.8%; Pred. No. 0.00026;
Matches 319; Conservative 210; Mismatches 608; Indels 557; Gaps 63;

alpha4(IV) collagen chains are arranged head-to-head on chromosome 2q36.";
 FEBS Lett. 424:11-16(1998).
 [3]
 RP SEQUENCE OF 1219-1690 FROM N.A.
 RC TISSUE-EYE; PubMed=8365481;
 RX MEDLINE=93374047; PubMed=8365481;
 RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;
 "cDNA isolation and partial gene structure of the human alpha 4(IV)
 collagen chain.";
 FEBS Lett. 330:122-128(1993).
 [4]
 RP SEQUENCE OF 1407-1507 FROM N.A.
 RX MEDLINE=93054733; PubMed=1429714;
 RA Kanagata Y., Mattei M.-G., Ninomiya Y.;
 "Isolation and sequencing of cDNAs and genomic DNAs encoding the
 alpha 4 chain of basement membrane collagen type IV and assignment of
 the gene to the distal long arm of human chromosome 2.";
 J. Biol. Chem. 267:23753-23758(1992).
 [5]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=9738662; PubMed=9195222;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 "The clinical spectrum of type IV collagen mutations.";
 Hum. Mutat. 9:477-499(1997).
 [6]
 RP VARIANT AS SER-1201.
 RX MEDLINE=95078927; PubMed=7987396;
 RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,
 Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,
 Smeets H.J.M., Reenders S.T.;
 "Identification of mutations in the alpha 3(IV) and alpha 4(IV)
 collagen genes in autosomal recessive Alport syndrome.";
 Nat. Genet. 8:77-82(1994).
 [7]
 RP VARIANT FBH GLU-897.
 RX MEDLINE=96379660; PubMed=8787673;
 RA Lemmink H.H., Nilleen W.N., Mochizuki T., Schroeder C.H.,
 Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;
 "Benign familial hematuria due to mutation of the type IV collagen
 alpha4 gene.";
 J. Clin. Invest. 98:1114-1118(1996).
 [8]
 RP VARIANTS AS, AND VARIANTS.
 RX MEDLINE=99011253; PubMed=9792860;
 RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,
 Cochat P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;
 "Determination of the genomic structure of the COL4A4 gene and of
 novel mutations causing autosomal recessive Alport syndrome.";
 Am. J. Hum. Genet. 63:1329-1340(1998).
 CC -!- FUNCTION: Type IV collagen is the major structural component of
 glomerular basement membranes (GBM), forming a 'chicken-wire'
 meshwork together with laminins, proteoglycans and entactin/
 nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
 alpha 6(IV), each of which can form a triple helix structure with
 2 other chains to generate type IV collagen network.
 CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
 CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
 colocalized and present only in basement membranes of kidney, eye,
 cochlea, lung and brain.
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 domain (NC1) at their C-terminus, frequent interruptions of the G-
 X-Y repeats in the long central triple-helical domain (which may
 cause flexibility in the triple helix), and a short N-terminal
 triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 are involved in inter- and intramolecular disulfide bonding. 12 of
 these, located in the NC1 domain, are conserved in all known type
 IV collagens.
 CC -!- DISEASE: Defects in COL4A4 are a cause of autosomal recessive

Alport syndrome (AS) [MIM:203780], an hereditary disorder
 characterized by progressive glomerulonephritis, renal failure,
 hematuria, ocular abnormalities and deafness. The recessive form
 occurs equally between males and females.
 -!- DISEASE: Defects in COL4A4 are a cause of familial benign
 hematuria (FBH) [MIM:141200] or thin basement membrane disease.
 FBH is characterized by persistent hematuria, an electron
 microscopically detectable thin glomerular basement membrane (GBM)
 and an autosomal dominant mode of inheritance. Renal function
 remains normal. In children, differentiation between FBH and AS
 can be difficult, because both disorders are manifested by
 persistent hematuria and thin GBM at that age.
 -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; X81053; CAA56943.1; -;
 EMBL; AB008496; BAA25065.1; -;
 EMBL; D17391; BAA04214.1; -;
 PIR; A55360; CGHU1B.
 DR Gene; HGNC:2206; COL4A4.
 DR MIM; 120131; -;
 DR MIM; 141200; -;
 DR MIM; 203780; -;
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR014442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 21.
 DR ProDom; PD000007; C1g_helix; 3.
 DR ProDom; PD003923; ProcollagnC4; 1.
 DR SMART; SM00111; C4; 2.
 KW Extracellular matrix; Connective tissue; Basement membrane; Repeat;
 KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
 KW Polymorphism; Alport syndrome.
 FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 1690 COLLAGEN ALPHA 4(IV) CHAIN.
 FT DOMAIN 39 64 7S DOMAIN.
 FT DOMAIN 65 1459 TRIPLE-HELICAL REGION.
 FT DOMAIN 1460 1690 NONHELICAL REGION (NC1).
 FT SITE 94 96 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 145 147 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 189 191 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 310 312 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 724 726 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 785 787 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 989 991 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1206 1207 CLEAVAGE (BY COLLAGENASE)
 (BY SIMILARITY).
 FT SITE 1212 1214 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 1480 1569 OR 1566 (BY SIMILARITY).
 FT DISULFID 1513 1566 OR 1569 (BY SIMILARITY).
 FT DISULFID 1525 1531 BY SIMILARITY.
 FT DISULFID 1588 1686 OR 1683 (BY SIMILARITY).
 FT DISULFID 1622 1683 OR 1686 (BY SIMILARITY).
 FT DISULFID 1634 1641 BY SIMILARITY.
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 569 569 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 441 446 Missing (in AS).
 FT VARIANT 545 545 G -> A (in dbSNP:1800516).
 FT VARIANT 570 570 E -> Q. /FTId=VAR_008149.
 FT VARIANT 897 897 G -> E (in FBH). /FTId=VAR_001912.
 FT VARIANT 931 931 A -> T.

```

FT FT 1004 1004 /FTID=VAR 008151.
FT FT L -> P (in dbSNP:1800517).
FT FT /FTID=VAR 008152.
FT FT G -> V (in AS).
FT FT /FTID=VAR 008153.
FT FT G -> S (in AS).
FT FT /FTID=VAR 001913.
FT FT P -> S.
FT FT /FTID=VAR 008154.
FT FT P -> L (in AS).
FT FT /FTID=VAR 008155.
FT FT LQ -> S (in REF. 3).
FT FT CONFLICT 1659 1660
FT FT SEQUENCE 1690 AA; 164095 MW; ELE72F283A72BAAE CRC64;

Query Match 2.6%; Score 339.5; DB 1; Length 1690;
Best Local Similarity 21.1%; Pred. No. 0.00032;
Matches 375; Conservative 107; Mismatches 617; Indels 681; Gaps 92;

QY 761 DTGQNGPK-PPATLGADGP-----PPGP-----PTPPRTSRAPTEPTPASEATGAPTPP 809
DB :|||:
DB ETGDVGPPGPGGLGRPGACAGMIGPPGPGFPGGLPGLPGEAGIPGRPDS-APGPKGP 438
QY 810 PAPPSPAPPPVPEKEEKEETAAPVEEGEQPPAAELAVDTGK-AEEPVKSECTE 868
DB :|||:
DB 439 GSGFGLPGAP-----GLOGLPSSVIYCSVGNPQPQGIKGVGP 476
QY 869 EAEEGPAKGDAAEAATGAALKAEKGGSGRATTAKSSGAPODSSSATCSADEYDE 928
DB :|||:
DB 477 PGRGPKGEGNEGLCACFPGMGPPGPPGLPGRGSGKDLGLPWLGTK-----526
QY 929 ASGGKNRLLSPRSLLTGTGPPRANASFOKPLDLKQLKORAAATIPP-----IQV 978
DB :|||:
DB 527 ---GD-----PGPPGAEGPPGLP-----GKGASGPPCNKAGKGMVVS 563
QY 979 TKVHEPPREDAAPTKPAAPPPQNLQPDSDAPOPSSSR-----GKSRPAPPADKEA 1034
DB :|||:
DB 564 VGHKGER-----GDDGPP-----GFGQGSRGDRDHAGKSGDPPGPDHED 606
QY 1035 FAAEAQKLPD-----PPCWTSGLPFPVPPPREVIKASHAPDPSAFSAPPQHP--L 1084
DB :|||:
DB 607 ATPGCKGFPGLPPGKAGVPCCGLGFPFGPE--RGHP-----GVPHGFGVG 654
QY 1085 PLGLH-----DTARPVLPRPTISNPPLISSAKHPSVLBRQIGAISQMSVQLHV 1135
DB :|||:
DB 655 PDLGKQKGDITISCNVTYFGRHGPFGFDGPP-----GPKGFPQGA--PGLSGS--702
QY 1136 PYSEHAKAPVGVMTGLPLPM-----DPKLAFFSGVKOEQLSPRCQAGPPSLGYPT 1188
DB :|||:
DB 703 --DGHKGRPTGTAEIIFGPFGRCDMDP-----GFGKEKS--SPVGPPGPGSPGV--752
QY 1189 AQEASVLRGTALGSV--PGGSITKGPSTRVPSDAITRGSITHGTTPADVLYKGTITRI 1246
DB :|||:
DB 753 NGQKIGPDPAFGHLGPPGKRGSLGVPQIGKP-----RGD--PGCF-----791
QY 1247 IGEDSPSL-----DRGREDLS-----PKGHVIYEKKG-----HVLSEGGMSV 1286
DB :|||:
DB 792 -GAEGPAGIPGFLGKPGKREGHAGFVGPPGPHCSERCERAGPIPGQGLPGYFG---846
QY 1287 TQCKEDGRSSGPPPHETAAPKRTYDMGGRVGRVRAISSASTEGLMGRAIPERHSPHHLK 1346
DB :|||:
DB 847 -----SPGAPGKGQP-----GDVGPP-GPAGMKGLPGLPGRGCAHPGCL--886
QY 1347 EOHHRISITGIPRISVYEAQEDYLREAKLLKREGTPPPPPSRDLTEAYKTQALGPLK 1406
DB :|||:
DB 887 -----PCIFGPFGD-----DGLFPPGPK-----GPRG 909
QY 1407 L-----KPAHEGLVATVKEAG-RS:HEI:PRELHRTPELPLAPRLKEGSTIQG 1454
DB :|||:
DB 910 LPGFPFGPCEKKGCAECPCAKGEPKKGMSGLGPD-----RGLGAKGAIG 957
QY 1455 TPLKYDTGASTGSKKHDRSLRSLIGSPGRTTTPPVHFLDVMADARALERACYEESLKSRPGT 1514
DB :|||:
DB 958 PP--GDEGEMAIISQK-----GTPGEPGPP-----GDDGFPGERGD 991

```

```

QY 1515 ASSSGGSIARGAPVIVVELGKXPROSPLTYEDHGAFFAGHLPRGSPVTWREPTPLQEGSL 1574
DB :|||:
DB 992 KGTFQMQRG-----ELGR-----YGP--GFHRGEPGEKGGPPGPPGPP 1032
QY 1575 SSSKASQDRKLTSTPREIAKSPHSTVPEHHHPHPIPSYEHLLRGVSGVDLYRSHIPLAFDP 1634
DB :|||:
DB 1033 GSTGI---RGFIGFP-----GLPDQGEPSGPG--PGFSGIDGARG-----1069
QY 1635 TSIPRGIPLDAAAAYILPRHLAPNPTYPHYLPYPPYLRGYPDPTALENRQTIINDYITTSQ 1694
DB :|||:
DB 1070 ---FKNGKDPAS-----HFCPPG-PKGEFGSPGCPG-----1097
QY 1695 MHNHTATAMAQADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVPPPTPGTATA 1754
DB :|||:
DB 1098 -HFG-----ASGEQLFGIQ-----GPRG-----SPGRPPGPG--1124
QY 1755 MDRLAYLPTAPQPPSSRSHSSPLSPGGPTHLTPTTTSSERERDRDRDREREREKSI 1814
DB :|||:
DB 1125 -----SSGP--PGCPGDHGMPLRGQPGEMGD---PGRGLQGDPGI 1161
QY 1815 LTSTTTVEHAPIWRPCTEQSSGSGSGGGG--GSSSRPASHSHAHQHSIPRTOALQ 1872
DB :|||:
DB 1162 -----PGPGIKGPGSPGLNGLKGLKQKTKGASGLHD--VGP-----1199
QY 1873 QRPVSLHNTGMKGIIITAVEPSKTVLSTSTSSPVRP-----AATFPATHCPGLGTLGV 1928
DB :|||:
DB 1200 --PGVIGILKG--ERGDGSPGI-----SPGPRGKKGPPGPGSGSGPGF--1242
QY 1929 YPTLMEPVLLPKAPRVARPERPRADTGHAFKAPPARSGLEPASPSKSGSEPRPLVPPV 1988
DB :|||:
DB 1243 -----PAGATGRAPKIDIPGPPGQG--PGPDGPRCAGP-----PGL 1280
QY 1989 SGHATARTPAKNLAPHHASPDPAPASASDPHREKTSKPSIOELELRSLYGHG---2045
DB :|||:
DB 1281 FGSVDLLRGEFGDC---GLPFPFGPPGPPGPP-----GYKGFPG 1316
QY 2046 ---SSYSPEGVE--PVSPVSSPSLTHDKGLPKHLELDKSHLEGLRKPQPGVKLGE 2099
DB :|||:
DB 1317 CDKDKGQKGMGPPGPGQPHGFPFGPEKGLPG-----PPGRKGTGLPG--1361
QY 2100 AAHLPHLRPLPESQSSSL---LOTAPGVKHQRVVTVAQHISEVITODYTRHHPQQLS 2156
DB :|||:
DB 1362 ---PRGEPGPADVDDCPRI PGLCAPGMRGPEGAMGL-----PGMR 1400
QY 2157 APLPAPLYSPGASC---PVLDRRPPSDLYLPPPHGAPARSPHSEGGKRSPEPNKTS 2213
DB :|||:
DB 1401 GP-----PGGCKGEPGLDGRRGVD-----GVP--GSPGPPGRKGT-----1435
QY 2214 VLGGGEDGIE-PVSPPEGMTEPG-----HSRS-----AVY 2242
DB :|||:
DB 1436 ---GEDGYGPGPGPGICDPQPKGFGPGYLLGGLLVHLSQTDQPTCLGNPRLWTGY 1491
QY 2243 PLLYRDEQTEPSR-----MGSKSPGNTSOPPAFFSKLTSNSAMVKSKEKKNKLNTHN 2298
DB :|||:
DB 1492 SLLYLEGQKAKHQNDLGLAGSCLPVPSTLFPAYCN-----IHQVCHYAQ 1535
QY 2299 RNEPEVNI SQPGTEIENMPALITCTGLMTYRSQAVQEHASTNMCLEAIRKALMGKYDQWE 2358
DB :|||:
DB 1536 RNDSTWLSAAP---LP-----MPLSEAIRPYVSRCAVCEAPOAVHVSQDQ---1583
QY 2359 ESPPLSANAFNPLNASASLFAAMPITAAADGRSDHTLTSFG 2398
DB :|||:
DB 1584 -SIPPCPQWRSIWIGYSF--LMHTGAGDQGGQALMSFG 1620

RESULT 67
TF20_HUMAN
ID TF20_HUMAN STANDARD; PRT; 1960 AA.
AC Q9UGU0; Q14528; Q13078; Q9H4M0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

DE Transcription factor 20 (Stromelysin 1 PDGF-responsive element-binding
 GN protein) (SPBR-binding protein) (Nuclear factor SPBP) (AR1).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), CHARACTERIZATION, AND
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=20568288; PubMed=10995766;
 RA Redal C., Sjoettem E., Johansen T.;
 RT "The nuclear factor SPBP contains different functional domains and
 RL stimulates the activity of various transcriptional activators.";
 RJ J. Biol. Chem. 275:40288-40300(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruekewich R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley N.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.C.T.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsey A.H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Suleston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Akawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Doxman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Rao Z.,
 RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edelman L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tilahun Y., Wright H.;
 RT "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 RN [3]
 RP SEQUENCE OF 245-1960 FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=97323006; PubMed=9179496;
 RA Ohara O., Nagase T., Ishikawa K.-I., Nakajima D., Ohira M.,
 RA Seki N., Nomura N.;
 RT "Construction and characterization of human brain cDNA libraries
 RT suitable for analysis of cDNA clones encoding relatively large
 RT proteins.";
 RL DNA Res. 4:53-59(1997).
 RN [4]
 RP SEQUENCE OF 1083-1938 FROM N.A. (ISOFORM 2).

RX MEDLINE=98399843; PubMed=9730594;
 RA Rajadhyakha A., Riviere M., Van Vooren P., Szpirer J., Szpirer C.,
 RA Babin J., Bina M.;
 RT "Assignment of AR1, transcription factor 20 (TCF20), to human
 RT chromosome 22q13.3 with somatic cell hybrids and in situ
 RT hybridization.";
 RL Cytogenet. Cell Genet. 81:176-177(1998).
 CC -!- FUNCTION: Transcriptional activator that binds to the regulatory
 CC region of MMP3 and thereby controls stromelysin expression. It
 CC stimulates the activity of various transcriptional activators such
 CC as JUN, SP1, PAX6 and ETS1, suggesting a function as a
 CC coactivator.
 CC -!- SUBUNIT: Homodimer (Probable). Interacts with RNF4 and JUN (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=09UGU0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=09UGU0-2; Sequence=VSP_003984, VSP_003985;
 CC -!- TISSUE SPECIFICITY: Expressed in most tissues, except in ovary and
 CC prostate. Isoform 1 is exclusively expressed in brain, heart and
 CC testis, and this form predominates in liver and kidney. Isoform 2
 CC predominates in lung.
 CC -!- DOMAIN: The atypical PHD domain functions as a negative modulator
 CC of cofactor binding (By similarity).
 CC -!- SIMILARITY: Contains 1 A-T hook DNA-binding repeat.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -!- CAUTION: Ref.4 sequence differs from that shown due to a
 CC frameshift in position 1932.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR001965; Znf_PHD.
 CC SMART; SM00249; PHD; 1.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein; Alternative splicing.
 CC DOMAIN 1170 1191 LEUCINE-ZIPPER.
 CC DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC A.T HOOK.
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC NUCLEAR LOCALIZATION SIGNAL
 CC (BY SIMILARITY).
 CC PHD-TYPE (ATYPICAL).
 CC POLY-ALA.
 CC POLY-GLN.
 CC SER-RICH.
 CC POLY-GLN.
 CC POLY-PRO.
 CC PPLPC -> VRLWR (in isoform 2).
 CC /FTid=VSP_003984.
 CC Missing (in isoform 2).
 CC /FTid=VSP_003985.
 CC -----
 CC EMBL; AY007595; AAC28930.1; -; -;
 CC EMBL; AL031346; CAB42440.1; ALT_INIT.
 CC EMBL; AL021878; -; NOT_ANNOTATED_CDS.
 CC EMBL; AB006630; BAA22961.1; -;
 CC EMBL; U19345; AAC36392.1; ALT_FRAME.
 CC Genew; HGNC:11631; TCF20.
 CC MIM; 603107; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003677; F:DNA binding; NAS.
 CC GO; GO:0003713; F:transcription co-activator activity; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR0019

QY 956 SPQKPLDKLKQKORAAAIPIQ---VTKVHPPREDAAP-----TKPAPPAPPPON 1004
DB 387 -----PATSMPELOQPPVTPVLELPGSATVPVLELPGPLSTPVPPELPGPAT 433
QY 1005 LQESDAPQOPQSPGRKSRPAPPADKEAFAEAQKLPDPPCWTSGLPFP---VPPR 1060
DB 434 AVPELPGP-----SVTPVQL-----SQELP-----GLPAPSMGLEPQ 467
QY 1061 EVIKASPHAPDPSAFSAFPGHPLPLGLHDTARPVLPRPTTINPPPLISSAKHPSVLER 1120
DB 468 EV-----PEPSMAQELFGPLV-----TAAVELPEQPAVTVAMELTEQPVTTTELEQ 515
QY 1121 QIGAIQCMQSVQLHVPYSHAKAP--VG--PVTMGLPLBMDPKKLAPFGVKGQELSPPRG 1176
DB 516 PV-----GMTVEHGHPEVTTATGLGQPEATWVLELPGQ-----VATTALELPG 562
QY 1177 Q-----AGPPSLGVPTAOEASVLRG-----TALGSVFGSITKGPSTRVPDSAITYRSI 1229
DB 563 QPSVTGVPELPGPLSATRALELSCQPVATCALELPGPLMAAG-----ALEFSGQS 612
QY 1230 THGTPADVLKGTITRIIGEDSPSLDRGREDLSLPGHVIYEGKGVLSYEGGMSVTQC 1289
DB 613 GAAGALELQPLATGVL--ELPGQGAPELPGQPVATVLE--ISVQSVVTTSEL 664
QY 1290 SKEDGRSSGPPHETAAPKRTYDMMEGR-----VGRAISSASIEGLMGRAIPPERHSPH 1344
DB 665 STMTVSQSLEVFSTTAL--ESINTVAQELPTVLGSETSVTCVDPLMA-----PESHILAS 718
QY 1345 LKEQHIRGSIT---QGPTRSVEAQEDVLRREKLLKREGTPPPPPPSRDL--TBAYKTQ 1400
DB 719 NTWETHILASNTWDSQMLASNTMSQ---MLASNTWDSQMLASNTMSQMLASNTMSQ 774
QY 1401 ALGP-----LKLPAHEGLVATVKEAGRSITHEPRELHRTPELPLAPPLKESITQ 1453
DB 775 MLATSSMDSQMLATSSMDSQMLATSS-----MDSQMLATSSMDS 813
QY 1454 GTPLKYDTGASTGSKKHVDVRLSIGSPGRTPPVHPDLVDADARALERACYEELKSRPG 1513
DB 814 QMLATSSMDSQMLATSSMDSQMLATSS-----TMSQMLATSSMDSQMLA--- 857
QY 1514 TASSSGGSITARGAPVIVELGKPROSLTYEDH-----CAPAGHLPGRSPVTWREPTPL 1569
DB 858 TSMSDSQMLASG-----TMSQMLASGTMDAQMLASGT--MDAQMLASST 900
QY 1570 QEGSLSSKASQDRKLTTPREIAKSPHSTVPEHHPHPISPVEHLLRGVSGVDLYR--SHI 1628
DB 901 QDSAMLGSKSPDYRLAQDPYRLAQDPYRL--GHPDYRLGHDAYL-----QODPYRLGHD 954
QY 1629 PLAFDPTSIPIRGIPLDAAAAYLPHLAPNP--TYPHLY--PPYLIRGYPTAALENROT 1684
DB 955 PYRLTPD-----PYRMGPRPYRIAPRSYRIAPRPYRLAPRPLMLASRRS 998
QY 1685 IINDYITSOQM-----HNTATAMAQRA-- 1707
DB 999 MMSVAAERSMMSYERSMMSYERSMMSMAERSMMSYERSMMSMAERSMMSMAERSM 1058
QY 1708 -----DMLRGLSPRESSIALNYAAGPGIIT--- 1732
DB 1059 MSAYERSMMSYERSMMSMAERSMMSMAERSMMSYSAADRSMMSSYSAADRSMMSSY 1118
QY 1733 -----DLSQVPHLPVLVPT-----PG 1749
DB 1119 TADRSMMMAADSYTDSYDTYTAVMVPLPPEPPTMPLPPEPPTMPLPPEPPE 1178
QY 1750 TPATAMDRLAY-----LPTAQPPSSRRSSPLSPGPGTHLTPTTTSSSERDR 1800
DB 1179 GPALPTEQSALTAENTWTEVPSLP---SEBSVQPEFPVVSQBSISEPSAVPTDYSVAS 1235
QY 1801 DREDRDRERKSILTS--TTTVEHAPINWPCTQSGSGSGSGSGSGSGSGSGSGSGSHAH 1858
DB 1236 D-----PSVLVSEAAVTPPEP---PEPESSITL-----TPVESAVVAEEH 1273
QY 1859 QHSPISPT-----QDALQORPSVLHNTGMKGITTAVERPSKPTVLRSTS----- 1902

DB 1274 EVVPERPVTMCVSTPAMSAEFTVL-----ASEPPVMSETAETFDMSRASGH 1320
QY 1903 -----TSSPVRPAATF-----PPATHCPLGG-----TLGGVY 1929
DB 1321 VASEVSTSLVPAVTTVPVLAESILEPPAMAAPSSAMAVLESSAVTVLESSTVTVLESST 1380
QY 1930 PTLMEP--VLLPKEAPRVARPERPRADTGHAFKAP--PARSGLEPPASSPSKSGEPRPLVPP 1987
DB 1381 VTVLEPSVTVVPEPPVVAEPD-----VVTIPVFWVSALEPSV-----PVLEP 1422
QY 1988 ---VSHATARTPAKNLAPHASPDPPAPASGD---PHREKTKQSPSIOSELELURL 2041
DB 1423 AVSVLQPSMIVSEBSVSQVSTVTVSEPAVTSEQTVIPEVAIESTPMILESSIMSSH 1482
QY 2042 GYHGSYSPPGEVPEVSP--VSPSLTHDKLPGKLEELDLSHLELGRPKPG----- 2092
DB 1483 VMKGINLS--SGDQNLAPFEGHQAIALHSGEPPHAE---HLKGDFFYESHGINIDLIN 1537
QY 2093 -----PVKLGGEEAHLPLRPLPESQPSQSSPLLOTAPGV----- 2126
DB 1538 NHLIAKEMEHTVCAAGTSPVGEIGEE---KILPSTKQRTVLDTPVGVSEADAGE 1591
QY 2127 -----KGHQRVTVLAQHIS-----EVTQDYTHHHQQOQSAPL 2159
DB 1592 TLSTGPFPALEPDATGTSKGIE--FTTASTLSLVNKYVDVLSLTTQD--TEHDMVISTP- 1647
QY 2160 PAPLYSPFGACFVLDLRPPSDLYLPPPDHG-----APARGSPHSEGGKSPSEP 2209
DB 1648 -----SGGSEADIEGPLPAKDHLDLPSNNLVSKDTEEPVKSQDOTLAALLSPKE 1700
QY 2210 NKTSVLGGEDGIEPVSPGCMTEPGHRS-----AVYPLLYRDGQOTEPSPRG--- 2258
DB 1701 S-----SGGEKEYPP--PPKETLPDGSFANIEDINEADLVRLPLPKDMERLTSRAGIEG 1754
QY 2259 -----SKSPGNTSOPP-----AFFSKLTESNSAMVSK--KOEI 2290
DB 1755 PLLASDVGRDRSAASPVVSSMPERASESSEKDDYEIVFKVVDTHEKSKKNRDKGEK 1814
QY 2291 NKKLINTNRN-----EPEYNISQOTEIFNMPAITGTGLMTVRSQAOVQEHASTNMGLBA 2344
DB 1815 EKDRDSSLRSRKSRSSEHKSRTSESRSRARKRSKSKHSRQTRSRSRSSRRSSRS 1874
QY 2345 IIRKALMGK---YDQWEEPS-----PLSANAFLNASAS--LP 2378
DB 1875 RSKSRGRSRSVSKERKSPKRSRKRKRKRKRKRKRKRKRKRKRKRKRKRKRKRKRKR 1934
QY 2379 AAMPITAADGRSDHTLTSFGGKG--AKVSGRPS-----SRKAKSPA-----PGLASGD 2425
DB 1935 SRRRRRSVGRRRRSFISPSRRSRTPSRRSRTPSRRSRTPSRRSRTPSRRSRTPSRRS 1992
QY 2426 RPP-----SVSVHSEGDNCNRTPLTNRV-----WEDRPSASGTP-- 2461
DB 1993 RTPSRRRSRVVRRRSFISFVRLR---RSRTPLRRFRSFRSFRSFRSFRSFRSFRS 2049
QY 2462 -----FPYNPLIMRLQAGVWASP-----PPGLPAGSGFLAG----- 2493
DB 2050 LTDLDKAQLELAKANAAMAKAGVPLPPLNKPAPPPTIEEKVAKKSGGATIEBELTEKC 2109
QY 2494 -----PHAWDEPKP 2504
DB 2110 KQTAQSKEDDDVIVNKPVHVSDEEBEP 2136

RESULT 70

MUSB HUMAN
ID MUSB HUMAN STANDARD; PRT; 5703 AA.
AC Q9HCB4; O00447; O00573; O14985; O15494; O95451; O14881;
AC Q99552; Q9UE28;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 5B precursor (Mucin 5 subtype B, tracheobronchial) (High

FT	CARBOHYD	930	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	1277	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	1293	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	1557	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	1775	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	2192	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	2721	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	3419	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	3948	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	4745	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	4901	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	4958	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	4965	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	4987	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5037	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5052	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5156	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5427	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5467	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5506	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5507	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5543	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5553	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5604	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5618	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	CARBOHYD	5662	N-LINKED (GLCNAC. . .)	(POTENTIAL)	.
FT	VARIANT	5137	T -> S (in dbSNP:2672788).		
FT	CONFLICT	34	G -> E (IN REF. 2).		
FT	CONFLICT	95	FPGLCN -> LPLCK (IN REF. 2).		
FT	CONFLICT	104	S -> C (IN REF. 2).		
FT	CONFLICT	142	E -> K (IN REF. 1).		
FT	CONFLICT	225	R -> S (IN REF. 2).		
FT	CONFLICT	330	PL -> T (IN REF. 2).		
FT	CONFLICT	337	E -> N (IN REF. 2).		
FT	CONFLICT	356	E -> K (IN REF. 2).		
FT	CONFLICT	362	G -> R (IN REF. 2).		
FT	CONFLICT	369	MISSING (IN REF. 2 AND 3).		
FT	CONFLICT	374	D -> N (IN REF. 2).		
FT	CONFLICT	393	RT -> TR (IN REF. 2).		
FT	CONFLICT	468	RK -> GR (IN REF. 2).		
FT	CONFLICT	512	L -> P (IN REF. 2).		
FT	CONFLICT	585	GAA -> AH (IN REF. 3).		
FT	CONFLICT	601	A -> S (IN REF. 3).		
FT	CONFLICT	628	DP -> RS (IN REF. 2).		
FT	CONFLICT	633	F -> L (IN REF. 2).		
FT	CONFLICT	676	A -> P (IN REF. 3).		
Query Match 2.5%; Score 335.5; DB 1; Length 5703;					
Best Local Similarity 19.5%; Pred. No. 0.0013;					
Matches 409; Conservative 203; Mismatches 835; Indels 649; Gaps 91;					
Qy	687	PAASEEAAFPVVEDEMEASVSGNEEVEAEALHASGNEVPRGECSPATVNSS	746		
Db	2149	PSSFTGTPFPVLTTATTAAATSSST	2195		
Qy	747	DTE-----SIPSPH-----	780		
Db	2196	THGRSLSPSPHVTCTAWTSATSGILGTTHTEPSTGTSHT---PAATTGTTQHSFAL	2252		
Qy	781	GPPTPPRTSRAPTEP-----TPASEATGATPP-----PAPPSPSAPPVVPK	824		
Db	2253	SSPHSPRTTESPPSGTTPGHTTATSRATATPSKTRTSLLPSOPT-SAPITTVT	2311		
Qy	825	EEKEETAAAPVEEGEQKPPAAEEL-----AVDTGKAEPVKSECTEEAEGPAK	877		
Db	2312	MGCEPQCAWSEWLDYSYMPGPGSGDDFTYSNIRAGGAVCEQPLGLECRAQAPGVPLR	2371		
Qy	878	KDAEAAEATAGALKAEKKG-----GSGRATTAKSSGAPQSDSS	918		
Db	2372	ELQGVVECSLDFGLVCNRRQGVKFMCFNVEIRVFCNCTPSTPATSTATPSGTPG	2431		
Qy	919	ATCSADEVDAEGGDKNRLSLPRPSLLTPTGDPFRANASPOKPLDLKOLKQRAAIPQV	978		

Db	2432	TTW-----ILITELTTATTTESTGSTATPSTL-----RTAPPPKULT	2469	
Qy	979	TKVHEPPREDAAFPKPAAPPQPNLOPESDAPQPGSPGRKSRSPAPADKEAPAAE	1038	
Db	2470	TTATTP---TVTSSKATFSSSPGTATALP---ALRSTATTATSVTPIPSSSLGT---	2519	
Qy	1039	AQKLPDGPCCWTSGLPFPVPREVIVKASPHADPSAF-----SYADPGH	1082	
Db	2520	-----TWIRLSQTTTPTATMSTATPSTPETAHTSVLTATATTGATGSAVATPS-	2569	
Qy	1083	PLPLGLHDTARPVL-----PREPTISNPPPLISAKHPSVLERQICAISSQGNMSV-	1131	
Db	2570	STPGTAHTKVPITTTTGTFTATFSSSGTALTTPVWISITTTTPT-----TRGSTVT	2620	
Qy	1132	QLHVPYSEHAKAPGVPTMGLPLPMDPKLAPSGVKQBSLPRGQAGPPESLGVPTAQE	1191	
Db	2621	PSSIPGTHTATVLTITTTTVA-----TGSMAFPSSSTSTGTPPSL-----TTTA	2666	
Qy	1192	ASVLRGTAIGSVPGSGSITKGI-PSTRVPSDALTIRGSIITHGTPADVLVYKGTITRIIGED	1250	
Db	2667	TTI---TATGTTNPSSTPGTRIPVLTATTATPAITSTVTPSSAL--GT-THTPPVP	2720	
Qy	1251	SPRLDRGREDSLPKGHVIEYEGKHVLSYEGGMSVTQCSKEDGRSSSGPPHETAAPKRT	1310	
Db	2721	NTTATTGRLSPSSPHTV--RTAWTSATSGTLGTHITE---PSTGTSHTPAATGT	2773	
Qy	1311	YDMMEGRVGRAISSASIEGLMGRAIP-PERHSPHULKQHHIRGSIQTGIPRYSVRAQED	1369	
Db	2774	TQ-----HSTPALSPHPSSTRTTESPPSGTTPFGH-----TTATSRTTATATPS	2818	
Qy	1370	YLREAKLLKREGTPPPPPSRDLTEA-----YKTAQLGLPLKPAHEG	1413	
Db	2819	KTITSILL-----PSSPTSABITTVTMGCPQCAWSEWLDYSVMPGFG---SGDFFD	2868	
Qy	1414	LVAIVKEAGRSIHEIPREELRHTPELAPR-----PLKEGSIQTGTPLYKYDTGASTT	1466	
Db	2869	YSNIRAAAGGAVCEQ-----LGLECRQAQPGVPLRE-LQWVECSLDFGLVCR	2917	
Qy	1467	GSKKHVDVRSILGSPGRTFFPVPHPLDVMDARALERACYEESLK-----SRGTAS	1516	
Db	2918	N-----REQVGK-----FKMCFNVEIRVFCNCYCHCPSTPATSS	2951	
Qy	1517	SSGGSIAAGAPVIVPELKGPRQSLTYEDHGAPFA--GHLPRGSPVTMRPTPLQEGSL	1574	
Db	2952	TATPSSSTPGTTLTQTAAATTTATGTAIPSSTPGTAPPKVLTSQATTP-----TA	3006	
Qy	1575	SSSKA---SODRKLSTPREIAKSPHSTVPEHHHPHPISPYEHLLRGVSGVDLYRSHIPLA	1631	
Db	3007	TSSKATSSSSPRTATLPLVTSTATKSTATSTFTPIPSSTL-----GTTGTSQNRPPHPMA	3061	
Qy	1632	FDPTSIPRGIFLDAAYVLPRLHAPNPTYPHLYPPYLIRGVDDTAALENRQTIINDYIT	1691	
Db	3062	TNST-----IHFSST-----PETT-----	3075	
Qy	1692	SOOMHNTATAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLPPTPGTP	1751	
Db	3076	-----HTSVLTTKATTTTATSSMTSPSS-----TPGIT	3104	
Qy	1752	-----ATAMDRLAYLPTAPQPFSSRHSSSLSPGPGTHLTPTTTSSSERDRDRRED	1805	
Db	3105	WILTELTATTAATTAALP-----HGTFFSSTPGTTLTTPSTTATTVTPTGSTATAS	3155	
Qy	1806	ROREREKS--ILTSTTTTVEHAPIWRPGTQSGSGSGSGSGSSRRSPASHSHAHQHSPI	1863	
Db	3156	STRATAGTLKVLSTATTP-----TVISRAFPSS--SPGTATPALPRLSTA-----	3200	
Qy	1864	SPRTDALOQRPVSLHNTGMKGIITAVEPSKPTVLRSTSTSPVPAATFPFATHCPLGG	1923	
Db	3201	TTPTATSVTAIPSSSLGTATWRLSQITTP-----TATMSTATPSSST-PETVHTSTVL	3251	
Qy	1924	TLDGVYPTLMEVLLPKAPRVARPERPADTGHAFKAPPARSG-----LEPASSP	1975	

Db 3252 TTTAT-TTRTSGVATPSSTPGTAHTTKVPTTTTTTGTATPSSPGTALTTPPVWISTTTTTP 3310
Qy 1976 -SKGSEPLVPPVSGH-ATIAATPAKNA PHHASPDPAPPAGASDPHREKTSKPPSI 2033
Db 3311 TTRGSTVTPSGIPGTHHTATVLTITTTTAVATGSMATPSSSTQTSPTSLTTAT----- 3365
Qy 2034 QELBRLSLGYHGSSVSP- --GVBPVSPV- ----SSPSLTHDKGLPKHLEBLDKSHLEGE 2085
Db 3366 ----TITATGSTTPNSSTPGTPIPPVLTATTATPAATSTVTPS--SALGTTH----- 3413
Qy 2086 LRPQGPVKLGGEAAHLPHLPBPESOPSSPILOTA--PGVKGHVQVTLAQHISEVI 2143
Db 3414 -TPVPV-----NTTATHGKSLPPSPHTVPTAWTSAATSGILG-----TTHITS-- 3456
Qy 2144 TQDTRHHQPQSLAPLAPLYFPFCASCPVLDLRPPSDLYLPPDPHGCAPARGSPHSRG 2203
Db 3457 -----PSTGTSHTPAATGTTPQSTPALSSHPSPRTTESPPGTTTPG--HTKGT 3506
Qy 2204 KR-----SPENKTSVL----- 2220
Db 3507 SRTTATATPSKTRTSTLLPSSPTSAPITTVVTGCEPOCANSEWLDYSYMPGPGSGDFD 3566
Qy 2221 -----CIE-----PVSPEGTEPECH--SRSAVYPLLVRDGEQ----- 2251
Db 3567 TYSNIRAAGVACQPLGLECRAQAPGVPLRELQGVVECSLDPLVCRNREQVKGKFKMC 3626
Qy 2252 -----TPSRMGSKSCGNTSGPPAFFSKLT-----ESNSAMVKSQKQ 2288
Db 3627 FNYEIRVECCNYGCHPSTPATSTSTGTTGTTWILTKLTATTATTESTGSTATPSTQ 3686
Qy 2289 -----EINKKLTHNRNPEYNIQGTETIFNMPAITGTLMTYRSQVQEHAS 2337
Db 3687 GPPAGTHVSTATTPTVTSKATP---FSFSG-----TATALPALRSTATTPTAT 3734
Qy 2338 TNWGLEAIRKALMGKYQWEESSPLSANAPNLNASALPAMPITAADGRSDHT--L 2394
Db 3735 S---FTAISSSLGTTWR-----LSQTTTPMATMSTATPSSPTETV-----HTSTVL 3779
Qy 2395 TSPGGGKAKYSGRPSSRKAKSPAGLAGSDRPPSVSS-----VHSEGDGNRRTPLTNRV 2449
Db 3780 TT-----TATTGATGCVATPSTPGTAHTTKVPTTTTTTGTVPSSPGTATP---PV 3831
Qy 2450 W---EDRPSAGSTPFP-----YNPLIMR-----LOAGVMASPPPPGLPAGSGP 2490
Db 3832 WISITTTTSSSTVTPSSIPGTHHTPVLTTTQPVATGSMATPSSSTQTSSTGTPP 3887
RESULT 71
CA1H HUMAN
ID CA1H HUMAN STANDARD; PRT; 1516 AA.
AC P39060; O9UK38; O9Y6Q7; O9Y6Q8;
DT 01-SEP-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN COL18A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98164096; PubMed=9503365;
RA Saarela J., Vlika R., Rehn M., Purmonen S., Pihlajaniemi T.;
RT "Complete primary structure of two variant forms of human type XVIII
collagen and tissue-specific differences in the expression of the
corresponding transcripts.";
RL Matrix Biol. 16:319-328(1998).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,

Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
Roentthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
Shintani A., Sasaki T., Nagamine K., Mitsuuma S., Antonarakis S.E.,
Minoshima S., Shimizu N., Nordsiek G., Hornisch K., Brandt P.,
Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
Lehrach H., Reinhardt R., Yagou M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN (3)
RP SEQUENCE OF 834-1516 FROM N.A.
RX MEDLINE=94245237; PubMed=8188291;
RA Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
Olsen B.R.;
RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen
and localization of the alpha 1(XVIII) collagen gene to mouse
chromosome 10 and human chromosome 21.";
RL Genomics 19:494-499(1994).
RN (4)
RP SEQUENCE OF 1334-1516 FROM N.A.
RC TissuePlacenta;
RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
RT "Cloning and expression of human endostatin gene in Escherichia
coli.";
RN Submitted (SEP-1999) to the ENBL/GenBank/DBAJ databases.
RN (5)
RP INVOLVEMENT IN KNOBLOCH SYNDROME.
RX MEDLINE=20400145; PubMed=10942434;
RA Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
Passos-Bueno M.R.;
RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
and tumor growth, plays a critical role in the maintenance of retinal
structure and in neural tube closure.";
RL Hum. Mol. Genet. 9:2051-2058(2000).
RN (6)
RP VARIANT ASN-1437.
RX MEDLINE=21518361; PubMed=11606364;
RA Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
Zorick T., Soares F., Camargo A., Oliveira E.S., di Loreto C.,
Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
for the development of prostatic adenocarcinoma.";
RL Cancer Res. 61:7375-7378(2001).
CC -!- FUNCTION: COL18A1 probably plays a major role in determining the
retinal structure as well as in the closure of the neural tube.
CC -!- FUNCTION: Endostatin potentially inhibits endothelial cell
proliferation and angiogenesis. May inhibit angiogenesis by
binding to the heparan sulphate proteoglycans involved in growth
factor signalling
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long; Synonyms=NC-493;
CC IsoId=P39060-1; Sequence=Displayed;
CC Name=Short; Synonyms=NC1-303;
CC IsoId=P39060-2; Sequence=VSP_001155, VSP_001156;
CC -!- TISSUE SPECIFICITY: Present in multiple organs with highest levels
in liver, lung and kidney.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- POLYMORPHISM: There is an association between a polymorphism in
position 1437 and prostate cancer. Heterozygous Asn-1437
individuals have a 2.5 times increased chance of developing
prostate cancer as compared with homozygous Asp-1437 individuals.
CC -!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KNO)
[MIM:267750]; an autosomal recessive disorder defined by the
occurrence of high myopia, vitreoretinal degeneration with retinal
detachment, macular abnormalities and occipital encephalocele.
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
INTERRUPTED HELICES (FACIT) FAMILY.
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; AF018081; AAC39658.1; -
 EMBL; AF018082; AAC39659.1; -
 EMBL; AL163302; CAB90482.1; -
 EMBL; L22548; AAA51864.1; -
 EMBL; AF184060; AAF01310.1; ALT_INIT.
 PDB; 1BNL; 02-DEC-98.
 GlycoSuiteDB; P39060; -
 Genew; HGNC:2195; COL18A1.
 MIM; 120328; -
 MIM; 267750; -
 GO; GO:0005581; C:collagen; TAS.
 GO; GO:0007397; P:histogenesis and organogenesis; TAS.
 GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
 GO; GO:0007601; P:vision; TAS.
 InterPro; IPR008161; C1g_helix.
 InterPro; IPR008160; Collagen.
 InterPro; IPR008985; ConA-like lec_gl.
 InterPro; IPR001791; Laminin_G.
 Pfam; PF01391; Collagen; 7.
 Pfam; PF02210; TSPN; 1.
 ProDom; PD000007; C1g_helix; 1.
 SMART; SM00282; LamG; 1.
 SMART; SM00210; TSPN; 1.
 Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
 Polymorphism; 3D-structure.
 SIGNAL 1 23
 FT CHAIN 24 1516
 FT CHAIN 1334 1516
 FT DOMAIN 221 409
 FT DOMAIN 410 516
 FT DOMAIN 517 550
 FT DOMAIN 551 560
 FT DOMAIN 561 640
 FT DOMAIN 641 664
 FT DOMAIN 665 786
 FT DOMAIN 787 809
 FT DOMAIN 810 892
 FT DOMAIN 893 906
 FT DOMAIN 907 948
 FT DOMAIN 949 961
 FT DOMAIN 962 1034
 FT DOMAIN 1035 1044
 FT DOMAIN 1045 1077
 FT DOMAIN 1078 1089
 FT DOMAIN 1090 1111
 FT DOMAIN 1112 1118
 FT DOMAIN 1119 1173
 FT DOMAIN 1174 1186
 FT DOMAIN 1187 1204
 FT DOMAIN 1205 1516
 FT CARBOHYD 68 68
 FT CARBOHYD 129 129
 FT CARBOHYD 164 164
 FT CARBOHYD 591 591
 FT CARBOHYD 1329 1329
 DISULFID 1366 1506
 DISULFID 1468 1498
 SITE 1095 1097
 VARSPLIC 1 180
 VARSPLIC 181 215

PRCPMPWPRRRRLDLVLAFLVLLGVRAASAE (in isoform Short).
 /FTid=VSP_001156.
 D -> N (increased risk of developing prostate cancer).
 /FTid=VAR_012709.
 F -> S (IN REF. 2).
 I -> V (IN REF. 2).
 V -> L (IN REF. 3).
 P -> R (IN REF. 3).
 P -> L (IN REF. 3).
 P -> L (IN REF. 3).
 A -> P (IN REF. 3).
 L -> K (IN REF. 3).
 P -> A (IN REF. 3).
 P -> P (IN REF. 3).
 G -> G (IN REF. 3).
 R -> G (IN REF. 3).
 A -> G (IN REF. 3).
 LR -> CG (IN REF. 3).
 R -> T (IN REF. 4).
 S -> Y (IN REF. 4).
 SQ SEQUENCE 1516 AA; 153840 MW; 3C70F29A4476EE76 CRC64;
 Query Match 2.5%; Score 335; DB 1; Length 1516;
 Best Local Similarity 22.8%; Pred. No. 0.0004;
 Matches 304; Conservative 89; Mismatches 519; Indels 420; Gaps 71;
 QY 675 KERNARRKKKAPAAAEAAFPVVEDEMEASGVSGNEEVEAEALHASGNVPRG 734
 DB 443 REETGAALKRLLPAP-----PVTTPPLAGSSSTEDSRSEVEQTVAISLGAQT--- 492
 QY 735 ECGSPATVNNSDTESIPSHTEAAKDTGNG-----PKPPATLGADGPP--PGPPTPPRR 788
 DB 493 -LFGSDSVSTWDSVTRPGGRVKEGGLKGKGPVGPVGPGRAGPGSPCLPGLPCP 551
 QY 789 TSRAPTEPT-PASEATGATPPAPSPAPPPVWVWKEEKEETAAPVVEESEEOKPPA 847
 DB 552 VS--PLGPAGPALQTVPGPGPPGPRDGT-----GRDGE-----PGDGEDKGP- 597
 QY 848 ABELAVDTGKAEEFVKSECTEAEAEFGPAKGDAAEAATAEGALKAEKKEGSGRATTAK 907
 DB 598 -----DTGPQGF-----GTPGDVGP-----KGDGDPGVGERGPPG 629
 QY 908 SSCAPQDSSS-----ATCSADEVDEAEGDKNRLSPRSLTPTTGDPRANASPKPLDL 963
 DB 630 PQGPPGPPGSPFRHDKLTFIDMEGSGFGDLEALRGPR-GFPGPPGPPGVLGPGE- 685
 QY 964 KOLKQRAAIP-PIQVTKVHEPRE--DAAPTKPAPPAP-----PP-----QNLQPEDSA 1011
 DB 686 GRFGVNSDVFGAGLPGV--PQREGPPGPGPLGPPGPPGREGPPGRTGQKSLGEAGA 743
 QY 1012 PQPGS-----SPRKSRSAPPA-----DKE----- 1033
 DB 744 PGHKGSKGAPGAPAGARGESGLAGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 803
 QY 1034 -AFAAEAQK-----LPGDPPCWTSGLP-----FP-VPPREVTKASHPAD 1071
 DB 804 TARSADGPPGPPGLPKRGDP--GVPLGPKAGEVADGIPGPPGLPGRG-AGPQGP 860
 QY 1072 PSFAYAPPGHPLPLGLHDTARFVLPFRPTISNPPPLISSAKHPSVLERQIGAIQGMVS 1131
 DB 861 GDRGSRGKGD- -GKDGVGQPLGPP- -GPPGVV-----YVSEQDSV----- 902
 QY 1132 QLHPVYSEHAKAPVG- -PVTMGLPLPMDPKLAPFGSVKQEQLS--PRGAGPPESLGVPT 1188
 DB 903 -LSVP-----GPEGRPGFAGPPGAPKGNL---GSKGERGSPGPKGEPGIFSPD 952
 QY 1189 AQEASVLRGTALGSPVPGSITKIGIPSTRVPSDAITYRGSITHGTTPADVLYKGTTRIIG 1248
 DB 953 -----GGALGPAQKA--KGEFGFRGPPG-----PYGRPG--YKGEI----- 985

Db 1527 SGVAGTGVGVFNHAFVSAHAFPGVRGACGATLGPSPLOAASASASSFOAPA 1586
Qy 1980 -----EPRLVPPVSGHATARTAKNLAPHASPD-----PPAPPASADPHREKTQ 2027
Db 1587 SVETRPDP 1646
Qy 2028 SKPFSIQEILRSIGYHGSSVSPGVSPVSPSLTHDKLPKHLKLELDKSHLEGELR 2087
Db 1647 S-----LASL-----PPNQAFLPPTSASL-----1666
Qy 2088 PKQGPVKLGEEAAHLPHLRPLPSQSPSSP-----LLQTPAGVKGHORVVTLAQHISEV 2142
Db 1667 --PPANASLSIKTSLPHKGRSPSTVHHQPLRLALAAQAPGIPQASATGPSAVVWSLG 1724
Qy 2143 ITQDYTRH 2150
Db 1725 MPPPYAAH 1732

RESULT 73
SMF1_HUMAN STANDARD; PRT; 1902 AA.
AC O14497; QUPZ1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE SWI/SNF-related, matrix-associated, actin-dependent regulator of
chromatin subfamily F member 1 (SWI-SNF complex protein p270) (B120).
GN SMARCF1 OR CLORF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE=20221560; PubMed=10757798;
RA Dallas P.B., Pacchione S., Wilsker D., Bowrin V., Kobayashi R.,
RA Moran E.;
RT "The human SWI-SNF complex protein p270 is an ARID family member with
non-sequence-specific DNA binding activity.";
RL Mol. Cell. Biol. 20:3137-3146(2000).
[2]
RN
RP SEQUENCE OF 1-1175 FROM N.A.
RX MEDLINE=98094256; PubMed=9434167;
RA Takeuchi T., Chen B.-K., Oiu Y., Sonobe H., Ohtsuki Y.;
RT "Molecular cloning and expression of a novel human cDNA containing CAG
repeats.";
RL Gene 204:71-77(1997).
[3]
RN
RP SEQUENCE OF 1-1132 FROM N.A.
RA Takeuchi T., Miesaki A.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS DNA NON-SPECIFICALLY.
CC -!- SUBUNIT: Part of the SWI-SNF complex.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: Highly expressed in spleen, thymus, prostate,
testis, ovary, small intestine, colon, and pbl, and at a much
lower level in heart, brain, placenta, lung, liver, skeletal
muscle, kidney, and pancreas.
CC -!- SIMILARITY: Contains 1 ARID domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
frameshift in position 1132.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AF265208; AAF75765.1; ALT_INIT.

DR EMBL; AB001895; BAA23269.1; ALT FRAME.
DR EMBL; AB024075; BAA83073.1; ALT SEQ.
DR EMBL; AB024059; BAA83073.1; JOINED.
DR EMBL; AB024060; BAA83073.1; JOINED.
DR EMBL; AB024061; BAA83073.1; JOINED.
DR EMBL; AB024062; BAA83073.1; JOINED.
DR EMBL; AB024063; BAA83073.1; JOINED.
DR EMBL; AB024064; BAA83073.1; JOINED.
DR EMBL; AB024065; BAA83073.1; JOINED.
DR EMBL; AB024066; BAA83073.1; JOINED.
DR EMBL; AB024067; BAA83073.1; JOINED.
DR EMBL; AB024069; BAA83073.1; JOINED.
DR EMBL; AB024068; BAA83073.1; JOINED.
DR EMBL; AB024070; BAA83073.1; JOINED.
DR EMBL; AB024071; BAA83073.1; JOINED.
DR EMBL; AB024072; BAA83073.1; JOINED.
DR EMBL; AB024073; BAA83073.1; JOINED.
DR EMBL; AB024074; BAA83073.1; JOINED.
DR GENE; HGNC:11110; SMARCF1.
DR MIN; 603024; -.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003677; F:DNA binding; NAS.
DR InterPro; IPR001606; ARID.
DR InterPro; IPR008938; ARM.
DR Pfam; PF01388; ARID; 1.
DR SMART; SM00501; BRIGHT; 1.
KW DNA-binding; Nuclear protein.
FT DOMAIN 96 99 POLY-GLN.
FT DOMAIN 178 184 POLY-GLN.
FT DOMAIN 615 618 POLY-SER.
FT DOMAIN 631 740 ARID.
FT DOMAIN 944 1021 GLN-RICH.
FT MUTAGEN 690 690 W-A: PARTIAL LOSS OF DNA-BINDING
ACTIVITY. WHEN A.A. 713 ALSO MUTATED,
COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 713 713 Y->A: PARTIAL LOSS OF DNA-BINDING
ACTIVITY. WHEN A.A. 690 ALSO MUTATED,
COMPLETE LOSS OF ACTIVITY.
FT CONFLICT 27 27 D->G (IN REF. 1).
FT CONFLICT 51 51 V->M (IN REF. 1).
FT CONFLICT 349 349 Q->S (IN REF. 2).
FT CONFLICT 368 368 G->GG (IN REF. 3).
FT CONFLICT 374 374 P->S (IN REF. 2).
FT CONFLICT 393 393 P->L (IN REF. 2).
FT CONFLICT 489 502 MANPPQVSGSMCP -> NGOYATSGWVRDVS (IN REF.
2).
FT CONFLICT 534 534 E->G (IN REF. 2).
FT CONFLICT 924 924 P->S (IN REF. 2).
FT CONFLICT 1016 1016 Q->L (IN REF. 2).
FT CONFLICT 1033 1033 Q->P (IN REF. 2).
SQ SEQUENCE 1902 AA; 205946 MW; 9B70AA46F81062EAA CRC64;
Query Match 2.5%; Score 332.5; DB 1; Length 1902;
Best Local Similarity 19.4%; Pred. No. 0.00059;
Matches 356; Conservative 202; Mismatches 664; Indels 609; Gaps 91;
Qy 728 GNEVPRGECGSPATVNNNS-----SDTESIPSPHTEAAKDTGQNGPKPATILGADGPP 779
Db 40 GSQTQ---RYPTMTVQGRASQAMGSLSYTQIIPYGGQSGYGGQGTQ---YNNQSPH 94
Qy 780 PGPTTPPRTSRAPTEPTPASEATGATPPAPPS-PSAPPV--VPKKEKEETAAAPP 836
Db 95 PQQQPP--YSQQPPSQTHAQSYPQSQPPQLQSSQPPYQSQSPPPHQQSPAPYP 152
Qy 837 VESEGEQ-----KPPAAELAVDTGKAEPPVKSECTEAEAGPAKGDAAEAATAGAL 891
Db 153 SQSSTTQQHPQSPQPPYQSQPPYQSQPPYQSQPPYQSQPPYQSQPPYQSQPPYQSQ 212
Qy 892 KAEKKEGSGRATTAKSSGAPQSDSATCSADVDVAEGG--DKNRLSPRPILLTPG 949
Db 213 RFPFPQLS-----QUSFGQASSAPMTSSKGGQEDNLSQSRPSL-----256
Qy 950 DPRANASPKPLDLKQLKQRAAAIPIQVTKVHEPPREDAP-----TKP-----AP 996

Db 257 -PDLGSGS-----IDDLPMGTGALSPGVSTSGISSQGEQSNPAQSPFSPHTSPHLPGLGIRG 311
Qy 997 PAPPP---PQNLOPESDAPQPGSSPRGKSRPAPPA-----DKEAFABAQKL 1042
Db 312 PPSFPVGSASVAQSGSLSPAAPV-GNQMPRPSPSGQSDIMHPMSQSSIAQDRGYM 370
Qy 1043 PGDP--PCWTSGLP-FPVPREVIKASPHA-----PDPSAFSVAPGHPL-PLGLHDTAR 1093
Db 371 QRPQWPQYSSPQFGSALSPRQSGGQIHTGMSVQONSMSGYSFGQGGYGPQGGY----426
Qy 1094 PVLPRPTTINPPPLISSAKHPSVLERQIGALS-QGMSVQLH-----VPYS-----EH 1140
Db 427 ---PRQNTNALP---NANYPISA-GMAGGINPMGAGGQHGQIGPIPYGYLPPGRMSH 477
Qy 1141 A---KAPVGPVTMGLPLMDPKKLAPFSGVKEQELSPRQAGPPPSLGVPTAQBASVLRG 1197
Db 478 ASGNRPYGNMANMP-----POVG-----SGMCPPPGGMNKT-QETAVAMH 519
Qy 1198 TALGSVPGSITKGPSTRPVSDSAITYRGSITHGTADVLKGTITRIIGEDSPSLDR 1257
Db 520 VAANSI-----QNRPPGVYNNQGMGTGPP---YQGINSMAGMINP-----Q 561
Qy 1258 GREDSLPKGHVYEGKKHVLSE-----GMSVTOCSKEDGRSSSGPPHETAAPK-----1308
Db 562 GPPYSM--GGTMANNASAGMAASPMMGLDVKLTPATKMNKADGTPKTESKSKSSST 619
Qy 1309 -----RTYDM-----MEGRVGRAISSAIEGLMGRAPERSPHIL-----KEQH 1349
Db 620 TTNEKITDYLGGEPERQWDRYLAFTTEKAMGNTLPAVGRKPLDLYLYSVKE--677
Qy 1350 HIRSGITO-----GIPRSYVEAQEDYLRREAKLLKREGTTP 1385
Db 678 --IGGLTQVKNKKNRELATNLVGTSSSAASSLKKYIQLYAF---ECKIERGEDPPP 732
Qy 1386 -----PPPSRDLTEAYKTOALGPLKLPKPAHEGLVATVKAGRSIHIP 1429
Db 733 DIFAAADSKSQPKTQPPSPA-----GSGSMQGPOTPQSTSSMAEGDGLKPTP 782
Qy 1430 REELRHTPELPLAPRLKEGSIQTGTPLYKDYGTASTGSKKHVRSLLGSGRTPFPVHP 1489
Db 783 AS-----THSQIPLPLPGMSRNSVIGIQAFNDGSDSTFOKN---SMTPNFGYO-PSMNT 834
Qy 1490 LDVMADARALERACTEE-----SLKSRPGT-----ASSGSGSIARGAP---VIVPELKG- 1535
Db 835 SDMMG-----RMSYEPNKDPYSGMRKAPGSDPFWMSGQGNCGMDPYSAAGFGLNV 888
Qy 1536 ---PRQSPITYEDHGAPFAGHLPRGSPVTMEPTPL-ORGLSSSSKASQDRKLTSTPRE 1591
Db 889 AMGPRQ-----HYPYGGPYDRVTRTFEGIGPEGNMSTG-APQNLMPSPNDS 933
Qy 1592 IAKSP-----HST-----VPEHHPHITSP-----YHLLRGV 1618
Db 934 GWSYSPRYPPQQQQQQQRHDSYGNQFSTQCTPSSGPPSPQIITMYQQQQNYKRPMDGT 993
Qy 1619 SGVDLYRSHIPLADPTSPIGRILDDAAAYLPRHLAPNPTYPHLYPPVILIRGYPTAA 1678
Db 994 YGPPAKRHEGEMYSVPYSTGGQGPQ-----QQLPPAPQP-----ASQOQAAQ 1037
Qy 1679 LENQTIINDYITSQMHNTATAMAQRADMLRGLSPRESSALNYAAGRGIDLSQVP 1738
Db 1038 PSPQDVYNOY---GNAYPATATAATERRP-----AGGPO-----NQFP 1073
Qy 1739 HLPVLVPTTGTATAMDRLAYLTPAPQPPFSRHSLSPLSPGGTTHLTKTPTTSSRER 1798
Db 1074 F-----QFGRDRV---SAPPGTNAQQNMPPQWGGP-----1101
Qy 1799 DRDRERDREREKSLTSTTVHEHPIWRPGTEQSGSGSGSGSGSGSSRRPASHAH 1858
Db 1102 -----IQASAEVAQGTWQ-GRNDMTYNTANRQSTGAPQGFAYHG---1142
Qy 1859 QHSPISPRTOALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSPVPAATFFPATH 1918

Db 1143 -----VNRTEMLHTDQANHEGWSHGTRQPPYGP-----SAPVPMTPPPPSNY 1189
Qy 1919 CPLGTTLDGVYPTLMEPVLLPKEAPRVARPERPRADTGHAFPAKPPARSGLPEASSPSK- 1977
Db 1190 QP-----PPSMQNHIPQVSSP-----APLPRP-----MENRTSPKS 1221
Qy 1978 -----GSEPRPLVPVSGHATARTPAKNLAPHUASP-----DPPAPPASASDPHREKTQ 2027
Db 1222 PFLHSGMMQKAGFPV-----PASHIAPAPVQPPMIRRDITFFPGSV-----BATQ 1267
Qy 2028 SKPFSIOELELSLGYHSGSSYSPRGVFPVSPVSPSLTHD-----2067
Db 1268 PVLKQRRLTKDQIG-----TPAWRVMSLSKGLLAETWALDTINILYDDNSTMTF 1321
Qy 2068 --KGLPKHLELD-----KSHLEGE-----LRP-----KQPGFVKLGGE 2099
Db 1322 NLSQPLGELLELLVEYFRRLCIEIFGILKEYEVDGPGQRTLLDFGRFSKVSPPAMEGEE 1381
Qy 2100 AAHL--PHL-----RPLPSQSPSSSPLLQTAPOV-----KGQHV 2132
Db 1382 EELLGPKLEEBEVEEVEDEIAFSGDKPKASENSEKLIKFKDKLPVKIVQKNDPFV 1441
Qy 2133 VTLAQHISEV-----ITQDYTRHHPQOLSAPLAPLYSPFGASCPLDLR 2177
Db 1442 VDCSDKLGRVQEPDPSGLLHWRIGGDTTEHIQTHFSKTELLP---SRPHAPC-----1491
Qy 2178 RPPSDLYLPPDHGAPARGSPHSEGGKRSRPNKTSVLGGEDGIBPVSPPE---GMT 2232
Db 1492 -PPA-----PRKVTVAEGTFTT-----DQSGPPPDGPPKRTATMD 1529
Qy 2233 EPGHRSRAVPLLYRDQ-EOTEPSPRMGSKSP 2262
Db 1530 DMLSTRST---JTEDGAKSSEAIKESKFP 1557
RESULT 74
SON MOUSE
ID SON MOUSE STANDARD: PRT; 2404 AA.
AC OQX47; OQCO12; OQCK6; OQXPS;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE SON protein.
GN SON.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=129/Sv;
RX MEDLINE=20408886; PubMed=10950926;
RA Wynn S.L., Fisher R.A., Pagel C., Price M., Liu Q.Y., Khan I.M.,
Zammit P., Dadrah K., Mazrani W., Kessling A., Lee J.S., Bulwela L.;
RT "Organization and conservation of the GART/SON/DONSON locus in mouse
and human genomes.";
RL Genomics 68:57-62(2000).
RN [2]
RP SEQUENCE OF 1-116 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus, Small intestine, and Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Glessi C., King B., Kochiwa H.,
Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustincin S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

Qy	1097	--PRPTTISNPPPLISSAKHPSVLE--RQIGAI SQGMSVQLHVPYSEHAKAPGCVPMGL	1151
Dd	500	EQPVTITTEFEQPVAMTTVEHFCHPEVITATGLIQ	541
Qy	1153	PLPMDPKKLAPFSGVKOQLSPROQ---AGPPESLGIVPTAQEASVLRG TALGS---VPG	1205
Dd	542	ELPGOP-----VATTALSELGQPSVTGVPELSGLPSATRALELSGQS VATGALELPG	593
Qy	1206	GSI TKGIPSTRVPSDSAITVGSITHCTPADLVLYKGTIIRIIGEDSFSRLDRGREDSL PK	1265
Dd	594	QLMATG-----ALEFGSQGAGALELLGOPLATGV	628
Qy	1266	GHVIYEGKGHVLSYEGGMSVTOCSKEDGRSSGPHETAAPKRTYDMMBGRVCGRAISSA	1325
Dd	629	----GQPG-----APELPQQPVATVA-----LEISVQSVVTT S	657
Qy	1326	SIEGLMGRAIPPERHSPhHLKEQHHRGSI TOGI PRSYVFAQEDYL RREAKLL KREGTTP	1385
Dd	658	ELSTM-----TVSQSLEVSTTALESY-----NTVA	683
Qy	1386	PPPSRDITEAYKTQALGPLKLPKAHEGLVATVKEAGRS THEIP REEL-RHTPELPI APR	1444
Dd	684	QELFTTLVGESTVITGVDP LMAQESHMLASN TMETHMLASN TWSOMLASNTWDSOMLAS	743
Qy	1445	PLKEGSTIOGTPLKYDTCASTTGSKKHDRVSLIGSPGRT PPVHPLDVMA-----DARALE	1500
Dd	744	N T W D S O M L A S T M D S O M L A S T M D S O M L A T S T M D S -----QMLATSMDSOMLA	792
Qy	1501	RACYEESLKSRP-----GTASSGGSIARGAPV I P E L G K P Q S P L T Y E D H G A P F A C H	1553
Dd	793	TSSMDSOMLATSSMDSOMLATSSMESOMLAGA-----MDSOMLAS--GTWD AQM	840
Qy	1554	LPRGSPVMREPTPRLOGESSUSSKASODKRLTSTPREIAKSPHSTVPEHHHPISPYE H	1613
Dd	841	LASGT-MDAQMLASSTQDSAMWGSKSDPVR LAQDPVRLAODPYRL--GHDPVRLGH DAY	897
Qy	1614	LLRGVSGVDLYR-SHIP--LAFOPTSI-PRGILPDAAAAYL PHLPAINTYPHYLPYYL	1669
Dd	898	RL-----QODPYRLGHDPYRLTPDPDYRVSPR-PYRIAPRSY--RIAPRP-----	938
Qy	1670	IRGYDPTAALENQTIIINDYITSOOM-----	1695
Dd	939	YRLAPRLMLASRRSMMSYA AERSMMSYSYERSMMSMPAERSMMSAYERSMMS	998
Qy	1696	--HNHTATAORA-----DMLRGLSPRESSL	1720
Dd	999	AYERSMMSMPAERSMMSGAYERSMMSGAYERSMMSGMADRSMMSMSSYSAADRSM	1058
Qy	1721	ALNYAAGPRGII-----DL SQVPHLPVLPVPPPTGTPATAM	1755
Dd	1059	MSSYSAADRSMMSSYTDORSMMMAADSYTDSYTDSYTEAVMVPLPPEEPPT-----	1110
Qy	1756	DLRAYLPTAQPPSSRHSSPLSPGGPTHLT KT TTTSSE RDRDRDRDRDRERERKSIL	1815
Dd	1111	----MP-LPPEEPPTPLPPEEPPEGPALST-----EQSAL	1143
Qy	1816	TSTTTTVEHAPIWRFGTEQSSGSGSGGGSSSRPASHSHAHQHSPI SPRTQDALQORP	1875
Dd	1144	TADNT-----WSTEVLSTGESLS-----QPEFPVSQSEISEPN AV-PANYSMS ESET	1190
Qy	1876	SVLNHTGMKGIITAVBP SKPTVLRSTSTSPVRPATFP-----PATHCPLGGTLDGV	1928
Dd	1191	SML---ASEAVMTVPBFAREP-ESSVTSAPVESAVVAEHMVPERPMTYMYSETTMS---	1243
Qy	1929	YPTLMPEVLLPKEAPVARPE-----RPRADTGHA F-----LAKP-----PARSGLE	1970
Dd	1244	---VEPAVLTS EAVIS ETS ETYDSNR P---SGHAISEV TW SLLEPAVTISQPAEDSL E	1296
Qy	1971	-----PASSPCKGE-----PRPLVPVPSGHATIA--RTPAKNL-APHHA	2007
Dd	1297	LPGMTVPASTMTWTTSPPVVAVTEIPPVAVPEPIMAVPELPTMAVVKTPAVAVPEPLVA	1356

RESULT 75

```

NFM_BOVIN
ID      NFM_BOVIN      STANDARD;      PRT;      810 AA.
AC      OT7788;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Neurofilament triplet M protein (160 kDa neurofilament protein)
DE      (Neurofilament medium polypeptide) (NF-M) (Fragment).
GN      NEF3 OR NFM OR NFM.
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC      Bovidae; Bovinae; Bos.
OC      NCBI_TaxID=9913;
RN      [1]
SEQUENCE FROM N.A.
RA      Hall W.D., Zhang L., Balin B.J., Sprinkle T.J., Spicer K.,
RA      Gearhart D.A.;
RT      "The bovine neurofilament M subunit has a novel set of KSP repeats
RT      normally restricted to NF-H.";
RL      Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: Neurofilaments usually contain three intermediate
CC      filament proteins: L, M, and H which are involved in the
CC      maintenance of neuronal caliber.
CC      -!- PTM: There are a number of repeats of the tripeptide K-S-P, NFM is
CC      phosphorylated on a number of the serines in this motif. It is
CC      thought that phosphorylation of NFM results in the formation of
CC      interfilament cross bridges that are important in the maintenance
CC      of axonal caliber (By similarity).
CC      -!- PTM: Phosphorylation seems to play a major role in the functioning
CC      of the larger neurofilament polypeptides (NF-M and NF-H), the
CC      levels of phosphorylation being altered developmentally and
CC      coincident with a change in the neurofilament function (By
CC      similarity).
CC      -!- SIMILARITY: Belongs to the intermediate filament family.

```


RN [7] SEQUENCE OF 1227-1289 FROM N.A.
RP MEDLINE=86104139; PubMed=3002437;
RA Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
RT "Isolation and partial characterization of genomic clones coding for
RT a human pro-alpha 1 (II) collagen chain and demonstration of
RL restriction fragment length polymorphism at the 3' end of the gene.";
RN Biochemistry 24:6343-6348(1985).
[8]
RN SEQUENCE OF 1176-1226 FROM N.A.
RP MEDLINE=84118798; PubMed=6320112;
RX Strom C.M., Upholt W.B.;
RA "Isolation and characterization of genomic clones corresponding to
RT the human type II procollagen gene.";
RL Nucleic Acids Res. 12:1025-1038(1984).
[9]
RN SEQUENCE OF 35-167 FROM N.A.
RP MEDLINE=89233138; PubMed=2714801;
RX Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
RA "Organization of the exons coding for pro alpha 1(II) collagen N-
RT propeptide confirms a distinct evolutionary history of this domain of
RT the fibrillar collagen genes.";
RL Genomics 4:438-441(1989).
[10]
RN REVIEW ON VARIANTS.
RP MEDLINE=91184577; PubMed=2010058;
RX Kuivaniemi H., Tromp G., Prockop D.J.;
RA "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
[11]
RN REVIEW ON VARIANTS.
RP MEDLINE=90036909; PubMed=2572591;
RX Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
RA Hollister D.W.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RN Hum. Mutat. 9:300-315(1997).
[12]
RN VARIANT SER-1074.
RP MEDLINE=90036909; PubMed=2572591;
RX Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
RA Hollister D.W.;
RT "Glycine to serine substitution in the triple helical domain of pro-
RT alpha 1 (II) collagen results in a lethal perinatal form of short-
RT limbed dwarfism.";
RL J. Biol. Chem. 264:18265-18267(1989).
[13]
RN VARIANT SEDC 1095-GLY--TYR-1330 DEL.
RP MEDLINE=89266907; PubMed=2543071;
RX Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;
RA "Identification of the molecular defect in a family with
RT spondyloepiphyseal dysplasia.";
RL Science 244:978-980(1989).
[14]
RN VARIANT OSTEOARTHRTIS CYS-650.
RP MEDLINE=90370826; PubMed=1975893;
RX Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
RA "Single base mutation in the type II procollagen gene (COL2A1) as a
RT cause of primary osteoarthritis associated with a mild
RT chondrodysplasia.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
[15]
RN VARIANT OI-IV VAL-717.
RP MEDLINE=91291136; PubMed=2064612;
RX Bateman J.F., Hannagan M., Chan D., Cole W.G.;
RA "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT valine substitution in osteogenesis imperfecta type IV. Detection of
RT the mutation and prenatal diagnosis by a chemical cleavage method.";
RL Biochem. J. 276:765-770(1991).
[16]
RN VARIANT OSTEOARTHRTIS CYS-650.
RP MEDLINE=91086471; PubMed=1985108;
RX

RA Eyre D.R., Weis M.A., Moskowitz R.W.;
RT "Cartilage expression of a type II collagen mutation in an inherited
RT form of osteoarthritis associated with a mild chondrodysplasia.";
RL J. Clin. Invest. 87:357-361(1991).
[17]
RN VARIANT HYPOCHONDROGENESIS GLU-984.
RP MEDLINE=93054548; PubMed=1429602;
RX Bogasert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
RA Cohn D.H., Eyre D.R.;
RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
RT 1(II) chain produces hypochondrogenesis.";
RL J. Biol. Chem. 267:22522-22526(1992).
[18]
RN VARIANT HYPOCHONDROGENESIS SER-705.
RX Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
RA Ramirez F., Vitale E., Lee B.;
RT "Characterization of a type II collagen gene (COL2A1) mutation
RT identified in cultured chondrocytes from human hypochondrogenesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
[19]
RN VARIANT WS-II ASP-198.
RP MEDLINE=93304428; PubMed=8317498;
RX Koerikko J., Ritvaniemi P., Haataja L., Kaeaeiaainen H.,
RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
RT for glycine alpha 1-67 and that causes cataracts and retinal
RT detachment: evidence for molecular heterogeneity in the Wagner
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
RL Am. J. Hum. Genet. 53:55-61(1993).
[20]
RN VARIANT SEMD CYS-840.
RP MEDLINE=93282819; PubMed=8507190;
RX Holderbaum D., Malemud C.J., Moskowitz R.W., Haqgi T.M.;
RA "Human cartilage from late stage familial osteoarthritis transcribes
RT type II collagen mRNA encoding a cysteine in position 519.";
RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
[21]
RN VARIANT SEMD ARG-285.
RP MEDLINE=93252400; PubMed=8486375;
RX Vakkula M., Ritvaniemi P., Vuorio A.F., Ala-Kokko L.,
RA Peltonen L.;
RT "A mutation in the amino-terminal end of the triple helix of type II
RT collagen causing severe osteochondrodysplasia.";
RL Genomics 16:282-285(1993).
[23]
RN VARIANT SEDC CYS-206.
RX MEDLINE=94063862; PubMed=8244341;
RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
RT family with an Arg75-->Cys mutation in the procollagen type II gene
RT (COL2A1).";
RL Hum. Genet. 92:499-505(1993).
[24]
RN VARIANT SEDC CYS-920.
RX MEDLINE=93315508; PubMed=8325895;
RA Chan D., Taylor T.K.F., Cole W.G.;
RT "Characterization of an arginine 789 to cysteine substitution in
RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
RT dysplasia.";
RL J. Biol. Chem. 268:15238-15245(1993).
[25]
RN VARIANT SEDC SER-1128.
RP MEDLINE=93140139; PubMed=8423604;
RX Cole W.G., Hall R.K., Rogers J.G.;

RESULT	77
TPR	HUMAN
ID	TPR_HUMAN
AC	P12270;
DT	01-OCT-1989 (Rel. 12, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	15-MAR-2004 (Rel. 43, Last annotation update)
DE	Nucleoprotein TPR.
GN	TPR.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Verte-
QC	Mammalia; Eutheria; Primates; Homo-
CC	Catarrhini; Homini; Hominidae; Homin-

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y00067; CAA8276.1; --
 CC PIR; A27864; A27864.
 CC Genew; HGNC:7734; NEF3.
 CC MIM; 162250; --
 CC GO; GO:0005883; C:neurofilament; TAS.
 CC GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
 CC InterPro; IPR006821; Filament_head.
 CC InterPro; IPR001684; IF.
 CC InterPro; IPR002957; Keratin I.
 CC Pfam; PF00038; filament; 1.
 CC Pfam; PF04732; filament head; 1.
 CC PRINTS; PR01248; TPB1KERATIN.
 CC PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 KW Glycoprotein.
 FT INIT_MET 0 0
 FT DOMAIN 1 103 HEAD.
 FT DOMAIN 104 411 ROD.
 FT DOMAIN 412 915 TAIL.
 FT DOMAIN 104 135 COIL 1A.
 FT DOMAIN 136 148 LINKER 1.
 FT DOMAIN 149 247 COIL 1B.
 FT DOMAIN 248 264 LINKER 12.
 FT DOMAIN 265 286 COIL 2A.
 FT DOMAIN 287 290 LINKER 2.
 FT DOMAIN 291 411 COIL 2B.
 FT DOMAIN 613 690 6 X 13 AA TANDEM REPEATS.
 FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT CARBOHYD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).
 SQ SEQUENCE 915 AA; 102317 MW; 5F5D3DF34C9D9E50 CRC64;

Query Match 2.5%; Score 329.5; DB 1; Length 915;
 Best Local Similarity 19.2%; Pred. No. 0.00037;
 Matches 200; Conservative 149; Mismatches 361; Indels 333; Gaps 40;
 2 SGSTQLVAQTRATPRYPHSLSPVQIARTHTDVLGLLEYQHSDRYASHLS-----54
 29 SPSSGFRSGNSRGSPSTVSSYKSLAPRLAYSSAMLSAESSLDFQSQSSLLNGSG 88
 55 PGSIITQPORRRPSSLSEFPQNGERSQELHLPESHYLPGLCKSEMFIKSRPRLELLP 114
 89 PGGDYKLSR-----SNEKEQLQGLNDRFAGYI-----EKVHYLEQQNKEIE--129
 115 DPLLRPSLLATQGPAGSEDLTKD-----RSLTGKLEPVSPFPPTDPELELVPPRLS 168
 130 -----AEIQALRQKQASHAQGLDAYDQEIHELATLEWVN-----HEKAQVQLDSHL- 177
 169 KEELIQNDVRDREITWVEQOISLKKQKQOLEEBAKPPPEKPVSPPP-----IESKH 223
 178 EEDIHRLKERFEELARLDDTEAATRALRKDIEEASLVKVELDKVQSLQDEVAFLRSNH 237
 224 RS-----LVQIYYDENRKKAAAHRIELGLGPQVLPYLPNQPSTPRHYENIKINQAM-- 276
 238 EEEVADLLAQI-----QASHITVE-----RKDYLTDLSTALKE 271
 277 -RKXLYFKRRNHARKWKQKQFYDQQLMEALEKKVKRIENNNRRRAKESKRYEYK 335
 272 IRSOLESHSDQNMHQAEEW-----FKCRVAKLTEAAEQNKEAI-----SSAKE-EIAEYRRQ 322
 336 -----QPFETRKQREIQERMQSVGQSGLSAARSHEVSEIIDGLSEGN-----384
 323 LQSKSIELESVRGKESLEROLSDEE-----RHNHDLSSYQDTTQOLENELRGT 372
 385 -----LEKQMR-----QLAVIPMLYDADQQRKIFNNMGLMADPMKVVYKDRQ 427

Db 373 KWEMARHUREYQDILLNVKMDIEIAAYRKL-----EGEETRFSTFAGSITGPLYTHRPPI 429
 QY 428 VMNWSSEQEKTFREKFWQHPKFNGLIASFLERKTVAE--CVLYYILTCKNENYKSLVRR 485
 Db 430 TISSKIQTKVEAPKLVQHK---FVEELIETKVEDEKSEMEALTAITEELAAAMKE 485
 QY 486 SYRRGSGQQQQQQQQQQQQQQQQPMRPSQEKDEK-EKEKEAEKEEKPEVENDKED 544
 Db 486 EKEAAEKEEPEEPEAEVAAKSPVKATAPEKVEEKEEKEEQQEVEEEDGAKSD 545
 QY 545 LLKEKTDGTDGNDKEKAVASKRKTANSQGRKGRITRSMANESSEAIIPQQAEL 604
 Db 546 QAEGSGSEKEG-SSEKEEGEQEGETEA-----EAGEEA-----579
 QY 605 ASMELNESSRWTEEMETAKKGLLLEHGRNWSAIARVSGKTVSQCKNFYNYKKRQNLDE 664
 Db 580 ---EAKKEKVEEKSEEVATK-----SPVEKKG-----SPVPSKPVVEKKGSPVPSK 658
 QY 785 PPRTSRAPLEPTPASEATGATPPAPPSPSPAPPV---VPKEEKEEETAAAPPE--E 839
 Db 659 PVEEKGKSPVSKSPVEEKAQSPVPSKAEVKGQEKVEEKEEKEVKEAPKEEKE 718
 QY 840 GEEKPPAAEELAVDTCKAEPEVKSCTEA-----870
 Db 719 KKEKPKDVPF-----KKKAEKSPVKEEAEVVTITKSVKHLKEKVEKPKLOQKEKE 774
 QY 871 -----EEGPAKG-----KDAEAAEATAEG--ALKAEKKEGGSGRATTAKSSGAPD 914
 Db 775 KAGGEGSBEESGDKAGKSRKEDIAVNGEVEGKEEVEQETKEKSGR-----EE 824
 QY 915 SDSATSCAD--EVDEAEGGDKN 935
 Db 825 EKGVTNGLDLSPADEKKGDKS 847
 RESULT 81
 T2D3 HUMAN
 ID T2D3_HUMAN STANDARD; PRT; 1083 AA.
 AC O00288; Q99721; Q9BR40; Q9BX42;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)
 DE (TAFII135) (TAFII-130) (TAFII1130).
 GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1] SEQUENCE FROM N.A.
 RP MEDLINE=97336072; PubMed=9192867;
 RX Mengus G., May M., Carre L., Chambon P., Davidson I.;
 RA "Human TAF(II)135 potentiates transcriptional activation by the AP-2s
 RT of the retinoic acid, vitamin D3, and thyroid hormone receptors in
 RT mammalian cells.";
 RL Genes Dev. 11:1381-1395 (1997).
 RN [2] SEQUENCE FROM N.A.
 RP MEDLINE=21638749; PubMed=11780052;
 RX Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavros J., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beare D.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Bridgman A.M., Brown A.J.,

RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.R., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.L., Hunt S.B., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leivaesalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A.C., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RT Rogers J.;
RL "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=97098442; PubMed=8942982;
RX Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RT "Molecular cloning and analysis of two subunits of the human TFIID
complex: hTAFII130 and hTAFII100.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
RN [4]
RP IDENTIFICATION IN THE TFIID-HAT COMPLEX WITH TAF5L; TAF6L; TADA3L;
RP SUPT3H; TAF2; TAF5; TRAP; GCN5L2 AND TAF10.
RX MEDLINE=93303588; PubMed=10373431;
RX Brand M., Yamamoto K., Staub A., Tora L.;
RT "Identification of TATA-binding protein-free TAFII-containing complex
subunits suggests a role in nucleosome acetylation and signal
transduction.";
RL J. Biol. Chem. 274:18285-18289(1999).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 870-918 IN COMPLEX WITH
RP TAF12.
RX MEDLINE=20063193; PubMed=10594036;
RX Gangloff Y.-G., Werten S., Romier C., Carre L., Poch O., Moras D.,
RA Davidson I.;
RT "The human TFIID components TAF(II)135 and TAF(II)120 and the yeast
RT SAGA components ADA1 and TAF(II)68 heterodimerize to form histone-like
RT pairs.";
RL Mol. Cell. Biol. 20:340-351(2000).
CC -!- FUNCTION: Makes part of TFIID is a multimeric protein complex that
CC plays a central role in mediating promoter responses to various
CC activators and repressors. Potentiates transcriptional activation
CC by the AP-2 of the retinoic acid, vitamin D3 and thyroid hormone.
CC -!- SUBUNIT: TFIID is composed of TATA binding protein (TBP) and a
CC number of TBP-associated factors (TAFs). Component of the TFIID-HAT
CC complex, at least composed of TAF5L, TAF6L, TADA3L, SUPT3H/SPT3,
CC TAF2/TAFII150, TAF4/TAFII135, TAF5/TAFII100, GCN5L2/GCN5, TAF10
CC and TRAP.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC -!- SIMILARITY: Contains 1 TAFH/NHR1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11354; CAA72189.1; -;
DR EMBL; AL137077; CAC36006.1; -;
DR EMBL; AL109911; CAC22312.2; -;

DR EMBL; U75308; AAC50901.1; -;
DR PDB; 1H30; 26-SEP-02.
DR TRANSPAC; T02328; -;
DR Genew; HGNC:11537; TAF4.
DR MIM; 601796; -;
DR GO; GO:0005669; C:transcription factor TFIID complex; TAS.
DR GO; GO:0016251; F:general RNA polymerase II transcription fac. .; TAS.
DR GO; GO:0005515; F:protein binding; TAS.
DR GO; GO:0003713; F:transcription co-activator activity; TAS.
DR InterPro; IPR007900; TAF4.
DR InterPro; IPR003894; TAF_hom.
DR Pfam; PF05236; TAF4; 1.
DR SMART; SM00549; TAFH; 1.
KW Transcription regulation; Nuclear protein; 3D-structure.
FT DOMAIN 590 682 TAFH/NHR1.
FT DOMAIN 39 42 POLY-HIS.
FT DOMAIN 52 57 POLY-ALA.
FT DOMAIN 98 101 POLY-GLY.
FT DOMAIN 142 148 POLY-ALA.
FT DOMAIN 268 275 POLY-PRO.
FT DOMAIN 331 337 POLY-ALA.
FT DOMAIN 680 683 POLY-PRO.
FT DOMAIN 808 813 POLY-ALA.
FT DOMAIN 828 831 POLY-ASP.
FT CONFLICT 105 117 PGPSRRPLVPA -> GRGLLQORGGRES
FT CONFLICT 136 136 A -> S (IN REF. 2).
FT CONFLICT 185 185 G -> GPG (IN REF. 2).
FT CONFLICT 233 264 MISSING (IN REF. 3).
FT CONFLICT 293 293 P -> L (IN REF. 3).
SQ SEQUENCE 1083 AA; 109943 MW; A645382752A0752 CRC64;
Query Match 2.5%; Score 329.5; DB 1; Length 1083;
Best Local Similarity 23.0%; Pred. No. 0.00044;
Matches 230; Conservative 94; Mismatches 336; Indels 341; Gaps 50;
Qy 701 EDEMEASGVSGNEEMVEEAEALH-----ASGNEVPRGSCSGPATVNN 744
Db 16 EVDEKVVSDLVGSLQSALAAASHHHLA PRTPEVRAAAAGALGNHVVS GPAGAAGAP 75
Qy 745 SSDTESTPSHTEAAKDTGQNGKPP-----ATLGADGPP-PGPPTPRRTSRAPTEPTA 799
Db 76 AAPAAGAPGA-----APEPPAGRAPGPGGPGPGPGAGPAGGAAQTLLGSAAL 117
Qy 800 SEATGAPTPPPAPPSAPPPVPKKEEETAAPVVEEGEEQKPPAAAEELAVDTGKAE 859
Db 118 -----GPAPPAKLRPP-----PEGSAGACAPV-----PAAAAVAAG----- 149
Qy 860 EPVKSECTEAEAGPAK--GKDAEABEA-----TAEGALKAEKKEGGSGRATTAKSGAP 912
Db 150 -----PEPAPAGPAKPAAGPAALAAARAGPGPGPGPGPGAGPAGGAAQTLLGSAAL 202
Qy 913 QDSSTATCSADEVDEAGGDKNLLSPRS-----LLTP-----TGDPRANASPKPLDL 963
Db 203 LNSHHA-----APAVSLVNGGPAALLPLPKPAAPGTVITQTPFVGAAPPAAPSPPAAP 259
Qy 964 KQLQRAAAITPIQV-TKVHEPPREDAAPTKAPAPPPQNLQPEDAPQPPGSSPRGK 1022
Db 260 AAPAAPAAPPPPPAPATLARPCHGHPAGPTTAAAVAPPPA-----AANGGSA--GA 309
Qy 1023 SRSPAPPADKEAFAAEAKLFGDPFCWTSGLPFFVP-----PREVIKASPHA----- 1069
Db 310 APAPAPAGGPA-CVSGOPGPG-----AAAAAPAGVKAESPKRVVQAAPPAQTLLAAS 362
Qy 1070 -----PDPSAFSVAPPCHPLPL-----GLH 1089
Db 363 GPASTAASMWIGPTMQGALSPAAVPPAPGPTGLPKGAAGAVTQSLSRTPATTSGIR 422
Qy 1090 DTARPV-----LPRPPTISNP-----PP--LISSAKHPSVLERQIGALSQGMVOLH 1134
Db 423 ATLTPVTVALPKLPQPP--QNTNTQNTQNLQPLQPMGVVRSNGQLLMIPOQALAQ-MQAQAH 479
Qy 1135 V-PYSEHAKAPVGPV-----TWGLP-LPMDPKKLAPPSGVKOE-----Q 1171

Query Match 2.5%; Score 325; DB 1: Length 1781;
Best Local Similarity 18.7%; Pred. No. 0.00096;
Matches 417; Conservative 273; Mismatches 727; Indels 810; Gaps 102;

QY 75 GNRSELHLRPPSHYLPGLG-----KSEMEFI-----ESKPRLELLPDPLL 118
DB 138 GQENRNIQIPSESNBELTQPTBSQANDIGFKVKFKVFKTKVKDKTE-NPDIVQ 196
QY 119 RPSPLATGOPAGSEDLTKDRSLTKLEVPSPSPHPTDPELELPPRLSKELQNMDR 178
DB 197 LUTVKDEGGAGAGDHQPSL-GAGEAASESEPKQSTE-----KPEETLKREQ 246
QY 179 VDREITWVQOISKLKKQQLLEEAAPKPEKVPSP--PIESKHSRLVQIIDEN-- 234
DB 247 SHAEISPPAESQAVECEKEEKEQEK--EFSKSAESPTSVTSETGSTFKFTQGA 304
QY 235 ---RKAABAAHRIEGLGPOVELPLYNQPSDTRQYHENTKINQAMKILILFKRNHARK 292
DB 305 GWRKKTST-----FRPKKEDV-----EASEKK-----KE 328
QY 293 QMKQPCORYDQIMEALEKKVERIE-----NNPRRAKSKVREYVEKOPPEIRKQRELOE 348
DB 329 QPEKVDTEDEKAVASEKLTAQAPQEPESAHEPRLSAEBYK--VELPSEEQVS- 385
QY 349 RMOQSVQORGSLMSAARSEHEVEIIDGLSE-----QENLEKQWOLAVIPP 397
DB 386 -----GSQCPSEKPAPLATEVFEDEKIEVHQEEVVAHVSTVEERTEEQTEV----- 434
QY 398 MLYDADOQRIKFINNGLMADPMKYKORQVNMMSBOEKETP-----REK 443
DB 435 ---ETAGSVPAELVGMDAEPQEAPEAKELVKL-----KETCVSGEDPTQCADLSPDEK 486
QY 444 FMOHPKNFGLIA-----SFLERKTVAECVLYVYLTKKNYKSLVRSYRREGSKSQOQQ 498
DB 487 VLSKPE-GVYSEVEMLSQEMKVGQSPKLKLTSTGLKLSKKQKGGKGGDEESGE 545
QY 499 QOQOQOQOQOQOQMPRSSOEKDEKEKEAEBEKEPEVENDKEDLLKERTDTSGEDN 558
DB 546 HTQ-----VPADSPDQEOEQGESSASSPEEPBITCLEGLAEVQ-----QDG 589
QY 559 DEKAVASKRKTANSQCRKRG-----RITSMANEANSEEAITPOQSAE 603
DB 590 EAEGATSDGEK-----KRGVTPWASFKNWTPKKRVRR--PSESDKDELDKVKSAT 641
QY 604 LASMELNESSRWTEEMETAKGLLEHGR-----NWGAIAARMVGSKTVSQCKN 651
DB 642 LSSTES-----TASMOEEMKGSVEEPKPEPKRVDTSVSWEALI-CVGSS----- 687
QY 652 FYFNVKQNLDELILQCHKKEKERNARRKKKAPAAASEEAAPPVV----- 700
DB 688 -----KKR-----ARR-----SSDEEGGPKAMGGDHQKADEAG 717
QY 701 EDERMEASGVGBEEMVEAEALHASNEVPRCEGCPATVNNSSDT-ESITSPHTEAA 759
DB 718 KDKETGTGILAGQE-----HDPGQSSSPQAGSTEGEGVSTWESFKRLVTPRK 769
QY 760 KDTQNGKPKPATLGAQPPPPPP-----TPPRRTSR-----APIE- 795
DB 770 KSKSLEEKESBDSIAGSGVEHSTPDTEPGKEESWVSISIKFTPGRRKKRPGKQEQAPVED 829
QY 796 --PTPASATGAPTPPAPSPSPAPPVVPKEE---KEEETAAPPVVEGEQ-KPPAAE 849
DB 830 AGPTGANE-----DDSDVPAVPUSEYDAVEREMEAQQAKGAQEQKAAAT 877
QY 850 ELAVDTGKAE-----EPVKSEC-----TEE-- 869
DB 878 EVSKELSESQVHMMAAVALDGTAAATIIERSPSWISASVTEPLEQVEAEALLTEVLE 937
QY 870 -----AEEGP-----AKGDAEAAEATAEGALKAKEKGGSGRATTAKSSGAPQSDSS 918
DB 938 REVIAEEPTVTEPLPENREARGDVTVSEALTEPE-----AVTAETAETAGPLGS--- 986
QY 919 ATCSADEVDEAEGDKNRLSPRLTPTGDPANASPOKPL-----DLKQLKQRAAAI 973

Db 987 -----BEGTEASAEETEMVSAVSQLTSDPTTTEATPVQEEGGVPDIEEQRRTOEV 1041
QY 974 PPIQVTKVHEPPREDAAPTKAPAPPQNLQOPESDAPQOPGSSPPGKSRSPAPPADKE 1033
DB 1042 LOVAEKVKESQ-----LPGTGGPEDVLQ-----PVQRAEAERPEQAEASGLUKE 1088
QY 1034 ---AFAAEAQKLPQDPPCMTSGLPFPVPPPREVIKASPHADPPSAFVAPPCHPLPLGLHD 1090
DB 1089 TDVVLKVDQAQEAKEP--FTQG-----KVGQT--TPSEFEKAP----- 1123
QY 1091 TARPVLPRPPTISNPPPLISSAKHPSVLERQIGAISQMSVQLHVPVSEHAKAPVGPVTM 1150
DB 1124 -----QVTESIESSE----- 1133
QY 1151 GLPLPMDPKLAPSGVGVKQBLSPRGQAGPPESIGLVPTAQEASVLRGTALGSGVGGSIK 1210
DB 1134 -LVTTQAEULA--GVKSQEMVME-QAIPDSEVETDSETDGTGTFVADFDAPGTQKD 1188
QY 1211 GIPSTRVPDSAITYRGSITHGTTPADVLVYKGTITRIIGEDSPSKLDRGRBDSLPKGHVY 1270
DB 1189 EIVEIHENEVA--SGTQSGGTEAEAV-----PAQKER--PPAPSSFVFQ 1229
QY 1271 EGKK-----GHVLSYEGGMSVTQCKEDGRSSSGPPHETAAPKRIYDMMEGVGRA 1321
DB 1230 EETKEQSKMEDTLEHTDKEVSVETVILSKTEGTQEA---DQYADEKTKOV----- 1277
QY 1322 ISSASIEGLMGRAIPPERHSPHLKEQHHRIGSITQICIPRSYVQAEQEDVLRRE----- 1374
DB 1278 ---PFFEGL-----EGSIDGTIVTSREKVTVALKKGEGTEAEC 1313
QY 1375 ---AKLKEGTPPPPPSRDL--TEAYKTOALGPLKLKPAHEGLVATVKBAGRSIHE 1427
DB 1314 KDDALELOSHAKSPSPVEREMVQVREKTEA-----EPH----- 1351
QY 1428 IPREELRHTPELPLAPRLKSGSITQCTPLKYDTGTASTTGSKKHVDVRSILIGSPGTFPV 1487
DB 1352 VNEEKLEHETAVTVSEVSKQLQTVNVI-----IDGAK--EVSSLEGSP--PPC 1398
QY 1488 HPLDVMDARALERACYEBSLSKSRPG---TASSGGSIGARGAPVIVPBLGKPROSLTY 1543
DB 1399 LGQE-----EAVCTKIQVOSSEASFTLTAABEEKVL-GETANILETGE-----TL 1443
QY 1544 EDHGAPFAGHLDRGSPVTWRETPRLQEGSLSSKASQDRKLTSTPRETAKSPHSTVPSH 1603
DB 1444 E-----PAGAHVLEE-----KSEKNED--FAAHPGEDA-----VP-- 1473
QY 1604 HPHIPSPYSHLLRGVSGVDLYRSHIPLAFDPTSIIPRGIPLDAAAAYLPRHLAPNPTVPH 1663
DB 1474 -----TGPDCAKSTPVIUSATT-KKGLSSD----- 1498
QY 1664 LYPPYLIRGYPDTALENQTIINDYITSOQMHNTATAMAQRAQMLRGLSPRESSLAIN 1723
DB 1499 -----LEG-EKTTSLKWKSDVEQVACQEVKVSVAIEDLEPENGIILETKSKLVQN 1551
QY 1724 YAAGPRGIIDLQVPHLPVLVPTPGTATAMDRLAYL-PTAPOPFSSRHSSSPLSPGCP 1782
DB 1552 -----IIQ-----TAVDQVFRTEETATEMLTSELQTO----- 1578
QY 1783 THLTPTTTSSSERDRDRDRDREREKSLTSTTTTVEHAPIWRPGTEOSSGSSGSSG 1842
DB 1579 AHVIRKADS-----QDAQETEKEGEBPQA-----SAQDETPTITSAKESESTAVGQA- 1625
QY 1843 GGGGSSSRPASHAHQHSPTSPRTQDALQORPSPVLHNTGMKIITATPEPSKPTVLRSTS 1902
DB 1626 -----HSDISKOMSEASE-----KTMVTEVE----- 1646
QY 1903 TSSPVRPAATPPATHCPGLGTLGCVYPTLMPEVLLPKE-----APRVARPERPRADTGH 1957
DB 1647 -----GST---VNDQLEEVVLPSEEGGAGTKSVFE---DQGH 1680
QY 1958 AFLAKPPARSGLPASPSPKSGSEPRPLVPVPSGHATARTPAKNLA PHHASPDPAPPAS 2017


```
DR EMBL; S57132; AAB25797.1; -.
DR PIR; S23810; S23810.
DR GenBank; HGNC:2193; COL16A1.
DR MIM; 120326; -.
DR GO; GO:0005597; C:collagen type XVI; TAS.
DR GO; GO:0007565; P:pregnancy; TAS.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like_1ec_gl.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Collagen; Hydroxylation;
FT SIGNAL 1 21
FT CHAIN 22 1603
FT DOMAIN 50 231
FT DOMAIN 232 374
FT DOMAIN 375 505
FT DOMAIN 506 520
FT DOMAIN 521 554
FT DOMAIN 555 571
FT DOMAIN 572 630
FT DOMAIN 631 651
FT DOMAIN 652 722
FT DOMAIN 723 737
FT DOMAIN 738 875
FT DOMAIN 876 886
FT DOMAIN 887 938
FT DOMAIN 939 972
FT DOMAIN 973 987
FT DOMAIN 988 1010
FT DOMAIN 1011 1432
FT DOMAIN 1433 1471
FT DOMAIN 1472 1577
FT DOMAIN 1578 1603
FT CONFLICT 418 420
FT CONFLICT 537 537
FT CONFLICT 1160 1160
FT CONFLICT 1163 1163
FT CONFLICT 1165 1165
SQ SEQUENCE 1603 AA; 157692 MW; E27D9A1D4E598A37 CRC64;

Query Match 2.4%; Score 323.5; DB 1; Length 1603;
Best Local Similarity 21.4%; Pred. No. 0.00098;
Matches 376; Conservative 143; Mismatches 632; Indels 609; Gaps 99;

QY 751 IPSHTEAAQDTGN-QPKPP-----ATLCAD---GPPPG----- 781
DB 179 VASVHVDCSSASSQPLGPRMRPVGHVFLGLDAEQKPVSFDLQVHVCDELVLVEG 238
QY 782 -----PPTPPRRTSRAP-----IBPTPASEAT-----GAPTPPPAPSPSAPPVV 822
DB 239 CCEILPAGCPETSARRDTQSNELIENPQSEKVVTRCFCLBEPQNSEVDAQLTGRIS 298
QY 823 PKEK-----BEETAA--APPVEESEQPNAEELAVDTGKAEPVVKSECTEAEAEPAK 876
DB 299 QKAERGAKVHQETAADECPCVHGARDSNVTLAPSGPKGKGERGLPGPPGSGKEG-AR 357
QY 877 GKDA-----BAAEATAEGALKAEKGGSGRATTAKSSGAPQDSATCSADEVDAEG 931
DB 358 GNDVIRISPAQLOAEGP-KGEKESGA-----LGPSGLPGSTG-----EKGQK 401
QY 932 GDNKRLSPRLSLTPTGDFRANASPKP-----LDLKLKORAAAIPIQVTKV 981
DB 1193 SGLPSPGPPGPIQGPAG--LDGLDKDGKGLRGDPGPPGMLGMPGPKGKTGHP 1250

402 GEK-----GDGINKGVFGKPGORDAPGEICVTGPKGQKGDGDFVGP----- 441
982 HEPPREDAAPTKPAPPAPPQNLQESDAPQOPGSGSPRGKSRSPAPPADKEAFABAQK 1041
442 -----EGLAGEPGPCLPGP-----FGILGPGTPG-DPGG---PPGPKDGSSGIFGKE 487
1042 LPGDPPCWTSGLPFPVPPR-EVIKASHPADPDSAFSAPPGHPLPLGLHDTARVLPVPRPP 1100
488 GPGGKP---GKPGVKGEKGDPCFVCTLPPEGFQNFVGLPCKPQPKG-----EPGDP 535
1101 TISNPPLISSAK-----HPSVLERQICAIQGSMSVOLHVPYSEHAKAPVGPVTMGLP-LP 1155
536 VRARGDPGIGIKGEKGPCL-----SCSSVVAQO-HLVSSSTGASGDVSGPGLGCLP 588
1156 MDPKKLAPFSGVKOEQLSPRGQAGPPESLGVPTAQEAASVLRGTALGSPVPGSITKGIPT 1215
589 -----GRAGVPGLKEK-GNFEAGPAGSPGPPGVPAGIKG-----AKGPFCE 632
1216 RVPDSDAITYRGSITHGTADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVIEYKKG 1275
633 PCPALS-----LQDGDV-RVVALPGPSG-EKGEPPG-PPFGGL-PGKQG 672
1276 HVLSYEGMSVTQCKSKEDGRSSSGPPHETAAKRTYDMGSRVGRGAISSASIEGLMGRAI 1335
673 K--AGERGL-----KQKGDAGNPGDPTGPTT-----GRPGLS-GEPCVQGPAG--- 714
1336 PPRHSPHLKEQHHRINGSITQ-----GIPRSVVEAQEDYLRREK-----LLKREGTP- 1384
715 -PKGEKDGCTACPSLQCTVTDMAGRQGPQKGEQGPGEVGRPKGPQGLPGVQPPG 773
1385 -----PPPPSRDLTEAYKTQALGPLKPAHEGLVATVKEAGRSIHEIPRESLRHTPE 1438
774 LKGVCQGEPPGPGVQ-----GP-QGEFAGPL-----PGIQGLP-----GPR 810
1439 LPLAPRLKEGSIQTGP-LKYDT-----GASTTGSKKHDVRLSLISPGSTFPVPLD 1491
811 GP--PGTGEKG-AQSGSPGVKGATPGVPPGASVSGPPGRD-----GQQQT----- 854
1492 VMADARALERACYEESLKSRPTASSSGSIARGAPVIVPELKGKPROSPITYEDHGAPFA 1551
855 -----GLRGTPEKGRG-----EKGEPP---GECSCPSQDLIFS---GMFGA 891
1552 GHL-----PRGSPVTWREP-----TPRLQEGSLSSSKASQ---DRKLTSTPREIAKS 1595
892 PGLWMGSSWQPGQPGPIPGPPGPGVGLQGVGNGLPGQGLTAELGSLPIE--- 947
1596 PHSTVPEHHHPHPSPEHILRGVSG--VDLYRSHIPLAFDPTSIPIRGIPLDAAAYLPR 1653
948 -----QHLKSIQCGCVQOQRAH----- 965
1654 HLAPNPTYPHLYPPYLI-----RGYPDTAALENRQTIINDYITTSQMHNTATAMAQR 1706
966 -----POYLVEKGEKDGQGLPGVPLDN-----CAQCFLSLER-----PR 1000
1707 ADMLRGLSPRESSALNAAAGPRGIIDLQSVPHLPVLV-----PP-----TPGTATAM 1755
1001 AEAARG-----DNSEGDPCGVSGPLPGPGLPGQGEBSGPPGMRGSPGPPG--- 1047
1756 DRLAYLPTAPQPPSSRRSSPLSPG--GPTHILKPTTTTSSSERDRDRDRDREREKS 1813
1048 -----PIGPPGPFGA-VGSFGLPGLQGERGLTGLTDKG----- 1080
1814 ILTSTTTVEHAPWRPCTEQQSSSG-----SSGGGGSSSSSRPASHAHQHP 1862
1081 -----EPGPPQPGVPGATPPGLPGIKGERGYTGSAGEKGPPOGSEGLPGPPGP 1132
1863 ISPRTO-----DALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSSVR-- 1908
1133 AGPRGERGPGQNSGEKDGQGFQPGTGTGSPGPGKVGSPGPPGPPGQAEKSGEIRGP 1192
1909 -----PAATFPATHCPGLGTLGVYPTLMPEVLL-----PKEAPRVARPERPADTGA 1958
1193 SGLPSPGPPGPIQGPAG--LDGLDKDGKGLRGDPGPPGMLGMPGPKGKTGHP 1250
```


QY 1503 ----CYEESLSRGTASS-----SGSIARGAPVIVPELG-----KPRQSPLT----- 1542
Db 838 GSGPSSPASTKSSSTKSSSTKSSGSSGYASSPAAGDPAPERRKKRPAARRPG 897
QY 1543 --YEDHGAPAGHLPGSPVMEPT-----PRLOEGL-----SSSKASQD 1582
Db 898 DGEDEG--LSGAALRGDGHGDEEDRGPRKRRLSLGLGPAPDPAPALLSSSSSED 955
QY 1583 RKLSTPREIAKSPHSTVPEHHPIPSYEHLLRGVSDLYRSHIPLAFDPTSIPIRGIP 1642
Db 956 DRU-----RRPLGMPPE-HAPDCGFRVPAG-----ETHTPR--P 988
QY 1643 LDAAAAYLP-----RHLPNPTYPHLYPPYLIIRGYPTAALENRQTIINDYITSQQMHNN 1698
Db 989 SEAAALAYCPEVARALVDQVPELWRPALTP--DPAALA-----HIAARR--- 1033
QY 1699 TATAMAORADMLRGLSPRESLALNYAAGPRGIDLSQVPHLPVLVPTPTGTATAMD-R 1757
Db 1034 -GAPLERRAAMRQIADPED-----VRVVVLYDPLPHEELCAEPAGAPPAWDPR 1083
QY 1758 LAVLTPAQPFSSR-----HSSPLSPGGPT--HLTKPTTSSERE----- 1797
Db 1084 RGLSALLAAFAHLCTPDGSHAWAGNWTGRPDIGRLNAQVLLSARDLGFAGAVEYLC 1143
QY 1798 -RDRDRERDREREKSLTSTTTVEHAPIRWPGTEQSSGSSGGGGSSRRPASHSH 1856
Db 1144 RLGAARER-----LIVDTIEDWPADGPA-----VG DYH 1172
QY 1857 AHQSPISPTQDAL-----QRPVSLHNTGKMG--IITAVEPS----- 1893
Db 1173 VYVRVLDPAAQCAVRWPGCRELRAAVLDSSIVGPAFCARVEASFARLHPGAELRLCLR 1232
QY 1894 KPTVLSTSTSSVPRPAATPPATHCPLGGLDGVVPTL--MEPVLLPKEAPVAREBP 1951
Db 1233 QDNVRYTVSTRAGRTVPVLPFRAY-----RORVLTVDGCKDMARQSRALGUGDPD-- 1284
QY 1952 RADTGHAFLAKPPARSGLPASP-----SKGSEPRP-----LVPPVS 1989
Db 1285 -FDGAAGFGRHRAANRWGLAPLPRFVSCGRRGLAELRPGGLPAELRAFCAALLEPDA 1343
QY 1990 GHATIAKPAKLNAPHASDPDPAPPASADPHREKTQSPFSIQLELESLG-----Y 2043
Db 1344 EAAPLVLTG-----AVAAAGAPPAVLWD-----FAPFETSVRAAAGGAVETH 1386
QY 2044 HGSSYSEGVPEVPVSSPSLTHDKGLPKHLELDKSHLEGLRKPQ-CPVKL 2096
Db 1387 RPAGASGAGAGGEDGDSVEIVGVRG-----GDGRPRGLGPIKV 1426

RESULT 87
MYS3 HUMAN
ID MYS3_HUMAN STANDARD; PRT; 2004 AA.
AC Q92794;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE MYST histone acetyltransferase 3 (Runt-related transcription factor
DE binding protein 2) (Monocytic leukemia zinc finger protein) (Zinc
DE finger protein 220).
GN MYST3 OR RUNXBP2 OR ZNF220 OR MOZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=96376968; PubMed=872817;
RA Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
RA Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,
RA Horsman D., Mitelman F., Volinia S., Watmore A.E., Houseman D.S.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";

RL Nat. Genet. 14:33-41(1996).
CC -!- FUNCTION: May represent a chromatin-associated acetyltransferase.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: Involved in acute myeloid leukemias through a chromosomal
CC translocation t(8;16)(p11;p13) involving MYST3 and CREBBP.
CC -!- SIMILARITY: Contains 2 PHD-type zinc fingers.
CC -!- SIMILARITY: Belongs to the MYST (SAS/MOZ) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; UA7742; AAC50662.1; --
CC Genew; HGNC:13013; MYST3.
CC MIM; 601408; --
CC GO; GO:0006323; P:DNA packaging; TAS.
CC InterPro; IPR005818; Histone H1/H5.
CC InterPro; IPR002717; MOZ_SAS.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF01853; MOZ_SAS; 1.
CC Pfam; PF00628; PHD; 2.
CC SMART; SM00526; H15; 1.
CC SMART; SM00249; PHD; 2.
CC PROSITE; PS01359; ZF_PHD_1; 1.
CC PROSITE; PS00016; ZF_PHD_2; 2.
CC Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
KW Nuclear protein.
FT ZN_FING 206 265 PHD-TYPE 1.
FT ZN_FING 259 313 PHD-TYPE 2.
FT DOMAIN 371 379 POLY-SER.
FT ZN_FING 538 560 C2HC-TYPE.
FT DOMAIN 788 801 POLY-GLU.
FT DOMAIN 989 995 POLY-GLU.
FT DOMAIN 1019 1026 POLY-ARG.
FT DOMAIN 1069 1078 POLY-GLU.
FT DOMAIN 1147 1150 POLY-LYS.
FT DOMAIN 1221 1242 GLU-RICH.
FT DOMAIN 1267 1302 GLU-RICH.
FT DOMAIN 1411 1414 POLY-SER.
FT DOMAIN 1593 1597 POLY-SER.
FT DOMAIN 1643 1704 GLN/PRO-RICH.
FT DOMAIN 1897 1977 MET-RICH.
FT SITE 1546 1547 BREAKPOINT FOR TRANSLOCATION TO FORM
FT MOZ-CBP.
SQ SEQUENCE 2004 AA; 225054 MW; 9FFBAC3792854BA CRC64;
Query Match 2.4%; Score 322.5; DB 1; Length 2004;
Best Local Similarity 19.8%; Pred. No. 0.0013;
Matches 282; Conservative 199; Mismatches 582; Indels 359; Gaps 62;
QY 16 EPRYPHSLSYVQIARTHTDVLGLBYQHHSRDYASHLSPGSIQ-----PQRR 64
Db 801 EPQCRERELEISVGKSVSHENKQDSYSEKKEPVMAFVSVSTRLSKQVLPHPDLSLANS 860
QY 65 RPSLLSEFQPGNERSQELHLRPESHSYLPGLKSEMEFTESKPRLELLPDPLRPSPL 124
Db 861 QPSRRGRWCRKRRKTKRFRFGDKDKSLLETSAPQEQYGCCEKGEATQEQYTESQ 920
QY 125 ATG--OPA--GSBDLTKDLSLTGKLEPVSPSPPTDPELELVPPRLSKKEELQNMDRV 180
Db 921 VASEEQPSQDGKPDLPK--RLSEGVP-----WRGQLKKSPEAL--KRLTEGSESLP 970
QY 181 REITMVEQIISKKKQQLKEEAAPPEPEKPVSPPIESKHSRLVQIITYDENRKAEEA 240
Db 971 RRYSEGDRAV--LRGFSSESEE---EEPESPRSSP----- 1002
QY 241 AHRILEGLGPQVELPLYNQPSDTRQVHENIKINQAMRKLLLYFKRNHARKQWKQK-FC 299
Db 1003 -----PILTKPT-----LKRKKPFLHRRRRVRKRKHNSVVT 1035

"Restricted homology between human alpha 1 type IV and other procollagen chains.";
Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653 (1985).
[8]
SEQUENCE OF 1-28 FROM N.A.
MEDLINE=89034231; PubMed=3182844;
RA Soaninen K., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
"The structural genes for alpha 1 and alpha 2 chains of human type IV
collagen are divergently encoded on opposite DNA strands and have an
overlapping promoter region.";
J. Biol. Chem. 263:17217-17220 (1988).
[9]
SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
TISSUE=Placenta;
MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
"The arrangement of intra- and intermolecular disulfide bonds in the
carboxyterminal, non-collagenous aggregation and cross-linking domain
of basement-membrane type IV collagen.";
Eur. J. Biochem. 176:617-624 (1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
alpha 6(IV), each of which can form a triple helix structure
with 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
CC -!- PTM: Lysines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M26576; AAA53098.1; JOINED.
DR EMBL; J04217; AAA53098.1; JOINED.
DR EMBL; M26550; AAA53098.1; JOINED.
DR EMBL; M26540; AAA53098.1; JOINED.
DR EMBL; M26542; AAA53098.1; JOINED.
DR EMBL; M26543; AAA53098.1; JOINED.
DR EMBL; M26544; AAA53098.1; JOINED.
DR EMBL; M26545; AAA53098.1; JOINED.
DR EMBL; M26546; AAA53098.1; JOINED.
DR EMBL; M26547; AAA53098.1; JOINED.
DR EMBL; M26537; AAA53098.1; JOINED.
DR EMBL; M26538; AAA53098.1; JOINED.
DR EMBL; M26548; AAA53098.1; JOINED.
DR EMBL; M26549; AAA53098.1; JOINED.
DR EMBL; M26551; AAA53098.1; JOINED.
DR EMBL; M26552; AAA53098.1; JOINED.
DR EMBL; M26553; AAA53098.1; JOINED.
DR EMBL; M26554; AAA53098.1; JOINED.
DR EMBL; M26555; AAA53098.1; JOINED.
DR EMBL; M26556; AAA53098.1; JOINED.
DR EMBL; M26557; AAA53098.1; JOINED.
DR EMBL; M26539; AAA53098.1; JOINED.
DR EMBL; M26558; AAA53098.1; JOINED.
DR EMBL; M26559; AAA53098.1; JOINED.
DR EMBL; M26560; AAA53098.1; JOINED.
DR EMBL; M26561; AAA53098.1; JOINED.
DR EMBL; M26562; AAA53098.1; JOINED.
DR EMBL; M26536; AAA53098.1; JOINED.
DR EMBL; M26563; AAA53098.1; JOINED.
DR EMBL; M26541; AAA53098.1; JOINED.
DR EMBL; M26564; AAA53098.1; JOINED.
DR EMBL; M26565; AAA53098.1; JOINED.
DR EMBL; M26566; AAA53098.1; JOINED.
DR EMBL; M26567; AAA53098.1; JOINED.
DR EMBL; M26568; AAA53098.1; JOINED.
DR EMBL; M26569; AAA53098.1; JOINED.
DR EMBL; M26570; AAA53098.1; JOINED.
DR EMBL; M26571; AAA53098.1; JOINED.
DR EMBL; M26572; AAA53098.1; JOINED.
DR EMBL; M26573; AAA53098.1; JOINED.
DR EMBL; M26574; AAA53098.1; JOINED.
DR EMBL; M26575; AAA53098.1; JOINED.
DR EMBL; Y00706; CAA68698.1; -.
DR EMBL; X05561; CAA23075.1; -.
DR EMBL; M10940; AAA52006.1; -.
DR EMBL; M11315; AAA52042.1; -.
DR PIR; S16876; CGHU4B.
DR Genew; HGNC:2202; COL4A1.
DR MIM; 120130; -.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g helix; 6.
DR ProDom; PD003923; Procollagn4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172
FT CHAIN 173 1669
FT DOMAIN 173 1440
FT DOMAIN 1441 1669
FT CARBOXD 126 126
FT DISULFID 1460 1551
FT DISULFID 1493 1548
FT DISULFID 1505 1511
FT DISULFID 1570 1665
FT DISULFID 1604 1662
FT DISULFID 1616 1622
FT CONFLICT 237 238
FT CONFLICT 241 241
FT CONFLICT 319 319
FT CONFLICT 719 719
FT CONFLICT 837 837
FT CONFLICT 842 842
FT CONFLICT 896 896
FT CONFLICT 904 904
FT CONFLICT 914 914
FT CONFLICT 998 998
FT CONFLICT 1010 1010
FT CONFLICT 1012 1012
FT CONFLICT 1358 1358
SQ SEQUENCE 1669 AA; 160611 MW; 3BBBA6DFFB9BA84 CRC64;
Query Match 2.4%; Score 320.5; DB 1; Length 1669;
Best Local Similarity 21.0%; Pred. No. 0.0013;
Matches 415; Conservative 133; Mismatches 750; Indels 677; Gaps 100;
Qy 728 GNEVPRGECGPGATVNNSSDTEIPSHTEAAK-DTGONG----PKPPATLGADGPPPGP 782
Db 64 GMOGEGP-QGPP--GQKGDTEGPGTGTGTGTRPGGASGYPGNGLPGIFGQDG-PPGP 119
Qy 783 PTPP-----RTSRAPTEPTPASEATGATP----- 808
Db 120 PGTPGCGNGTGERGPGPLPGPLPGFAGNPGPPGLPGMKDGPGEILGHVPMLLKGERGPPG 179

QY 809 -----PPAPSPSAPPVW-----PKEEKEETAAP 835
Db 180 IPTGPPGGLPGLGVPPGPGTGPFGPPGPGGKGQWGLSFGQPKGDKDQGVSGP 239
QY 836 PVEEGEQKPPAAELAVDTKAEAPVKSECTEABEG-----PAKGDAAEAATAGAL 891
Db 240 PGVPGQAQ-----VQE-----KGFATKGEKQKGFPGQMPGVCEKEGPKPGPRG-- 287
QY 892 KAEKEGSGRATTAKSGAPQSDSSATCSADEVDEAGGDKNRLSP-RPSLLTPTGD 950
Db 288 -----KPKGDGKGEKSGSPGPGPGYPLIGR-----QGQGEKGEAGPPGPGIYVIGT- 338
QY 951 PRANASPKPLDLKQLKQRAAAIPPIQVTKVHEPPREDAAPTKAPAPPAPPPQNQLQPSD 1010
Db 339 -----PLGEK-----GERGYPTGPRGEPGPGKPGFLPG 368
QY 1011 APOQSSPRGKRSAPPADK-----EAPAAEAQKLPDPPFCWTSGLPFPVPPREVIKA 1065
Db 369 QPGPPGLPVGQAGAPGPFGEKGEKGRGFPGTS-LPG--PSGRDGLPG----- 415
QY 1066 SPHAPDPSAFSAPPCHPLPLGLHDTARVLPRLPPTISNPPPL-----ISSAKH 1114
Db 416 -PGSP-----GPPQP-----GVTNGIVECQPGPDGPGPGIQQPGFISGEGKQG 464
QY 1115 PSVLEROIGAI-----SQMSVQLHVYSEHAKA-----PVGPTVMGLPLPMDPKKLAPFS 1165
Db 465 ESCLICIDIGYRPPGPPGPGPGIIGFPQPGAKGDRGLPGRDGVAGVPGPGQTGGLIQP 524
QY 1166 GVKQE-----QLSPRQAGPESIGVPTAQASVLRGTALGSPVGGSIITKIGPSTRVPS 1219
Db 525 GAKGEPEGYFDLRUKGDKDGFPGQPG-----MPGRA-GS-PGRDGHPLCPKPGK-GS 575
QY 1220 DSAITVRSITHTPADVLYKGTITRIIGEDSPSLDRGREDLSPKGVHVE-----GKK 1274
Db 576 PGSVGLKGE--RGPPGVGFGS-----RGDTGPPGPGGPGAGPGDK 617
QY 1275 GHVLSYEGMSVTCSEKEDGRSSGPPHETAAPKRTYDMMEGRVRAISSASIEGLMG-- 1332
Db 618 G-----QAGFPFGPSGLPGPKGEP-----GKIVPLPGPGASGLPGSP 657
QY 1333 -----RAIPERHSPHLKEQHIGRSITO-----GIPRSYVBAQEDYLREAKLKRE 1391
Db 658 GFPGQGDGFGTGRGFLGPE-----KGAVQPGIIGFPFGPGKGVNDGLPGD-----MGPP 710
QY 1382 GTTPPPP-----PSRDLTEAYTKQ-----ALGPLKLPALHEGLVATVKEAGR-SIHEIPREE- 1432
Db 711 GTPGPRPGNGLPGNPGVQKGEKPGVGLPKGLPLGPGIIGTPEGKSGIIGVPGVGEH 770
QY 1433 -----LRHTRELPLAPRLKEGSI-TOGTPKVDYDTGASTTGSK-----KHVRSILIG 1478
Db 771 AIGPPGLQIGRGPGLP-----GSVSGPGVP-----GIGPPGARGPPGQGGPGLSG 820
QY 1479 SPG-----RTFPVPHPLDVMA-----DAPALERACVYESLSRPGTASSSGSITARGAPVI 1529
Db 821 PPGIKERKGFPGGLDMPGPKGDKAGQLPGITGQSLGCLPLGQOQAPGIPGFGPSKGE 880
QY 1530 VPBLGKPRQ-----PLTYEDHGAFFAGHLPRGSPVTWRE-----PTP 1567
Db 881 MGVMGTQPGQSGPGVAGAPLPGKGDHGFPGSSG-PRGDFGLKGDGVDVGLPKPGKPSMD 939
QY 1568 LROEGSLSSSSKASQDRK-----LTSTPREIAK-----SPHSTVPEHHH--- 1606
Db 940 KVDGMSKMGKGDQGEKQIGPIGEKSGRDPGTGPGVPGKQAGQPGQPGKDPGDISG 999
QY 1607 -----PISPEHLLRGVSGVDLYRSHIPLAFDPTSIIPRGIPLDAAA 1647
Db 1000 TPAGAPLPGPKSGVGMGLPGTPE-----KGVPGIP-----GPGQSP-GLPGDKGA 1045
QY 1648 AYLPRHLAPNPTPHLYPPVLYRNGYDDTALENRTQIINDYITSQMHNTATANAQRA 1707
Db 1046 KGEKQAGPPGIGIPLG-----RG-----EKG 1067

QY 1708 DM-LRGL--SPRESSLALNYAAGPRGIIDLSOVPHLPVL-----VPTEGTPTATMD 1756
Db 1068 DOGIAGRPGSPGK-----GEKSGIGIPCMGPGSLKXGSPGVYPGSLPGBKGD 1119
QY 1757 RLAYLPTAPQ-PFSSRIUSSPLSPGGPHTLTKPTTTSSSERDRDRDRDREREKSIL 1815
Db 1120 K--GLPGLDGIPIGVKGEAGLFGTGF-----PTGPAGQKGEPCSGDGI PGSAKEGPGGL 1170
QY 1816 TSTTTVEHAPIWRPGTQBQSSGSSG-----GGGSSSRPASHSHAHQHSPTISPTODA 1870
Db 1171 PGRGF-----PGFPAKGDGKSGKEVGFPGLAGSPGIPGSGKEGFGWPGPGQPG 1222
QY 1871 LQOPSPVLHNT-----GMKGIITAVEPSKPTVLRSTSTSSVPRPAATPPATHCPGLG 1923
Db 1223 LPSFG--HATEGPKGRPGQ-----QPLGL-----PGMPGPG--LPG 1260
QY 1924 TLDGVYPTLMEPVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEPASSPKSGSEPRP 1983
Db 1261 -IDGVKDKGNP-----GWFGAPGVPGPKDGP--FGMPGI--GGSPGITSKGDMPGP 1310
QY 1984 LVPPVSGHATARTP-----AKNLAPHHASPPDPAPPASASDPHREKTQSKP 2030
Db 1311 GVPGFQPKGLPGLQIGIKGQDQGVGAKGL-----PCPPGPG-----P 1351
QY 2031 FSTOELELSLGYHGSYSPEGVEPVSSPSLTHDKGLPKHLELDKSHLEGELELPK- 2089
Db 1352 YDIKGE--PGLPG-----PEG-----PPGLKGLGLP-----GPKG 1381
QY 2090 QPGFVKLGEEAHLHLRPLPESQSSSPLLQTAGVKGHQVVTLLAQHI SEVITQDYTR 2149
Db 1382 QQGVTVLGV-----IP--GPGIIGFGDAPGQKG----- 1408
QY 2150 HHQQLSAPLADPAPLYFPGASCVDLLRPPSDLYLPPDHGAPARGSPHSGKRSPEF 2209
Db 1409 -----EMGAPGTGPRGPGP-----PGPD-----GLPGSGPPTPSV 1442
QY 2210 NKTSVLGGEDGIEPVSPPEGMTEPGHSSAVYPLLYRDG-EQTEPSRMGSKSP--GNVS 2266
Db 1443 DHGFLVTRISQIIDDQCSGKILYHG-----YSLLYVQGNERAHQDGLTAGSCLRKFS 1498
QY 2267 QPPAFPSKLTESAMVSKSQEINKLNTNRNPEYNISQPGTEIFNMPAITGTGLMT 2326
Db 1499 TWPFLPC-----NINNVCFASRNDYSYWLSTPEPMPSMAITGENIRP 1543
QY 2327 YRSQAQVEHASTNMGLEALIIRKALMGKYDOWEESPLSANAPNPLNASASLPAAMPITAA 2386
Db 1544 FISRCVACBAP-----MVAVHSQTIQIPP-CPSGWSLWIGYSFVM----- 1585
QY 2387 DGRSDHTLTPGGGGKAKVSGRPSRRKAKSPAPGLASGD-RPPSVSVSHSEDCN 2440
Db 1586 -----HTSAGAEAGSGOALAS-----PGSCLEFRSAPFIECHGRGTCN 1623

RESULT 89
CA21-CHICK
ID CA21-CHICK STANDARD; PRT; 1362 AA.
AC P02467; P87491; P87492; Q90758; Q90792; Q90795; Q90797; Q92014;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(I) chain precursor (fragments).
GN COL1A2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-245; 262-448 AND 466-1362 FROM N.A.
RX MEDLINE=86185168; PubMed=3868961;
RA Boedtker H., Finer M., Aho S.;
RT "The structure of the chicken alpha 2 collagen gene.";
RL Ann. N.Y. Acad. Sci. 460:85-116(1985).

101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000


```
DB 1151 KNPAR-----TCRDLRLSHPWSSGFYWDNPQGCTA 1182
RESULT 90
CA2B_MOUSE
ID CA2B_MOUSE STANDARD; PRT; 1736 AA.
AC Q64739; Q61432; Q921W0;
DT 15-JUL-1998 (Rel. 36, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 2(XI) chain precursor.
GN COL1A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 7).
RC STRAIN=129/SVJ;
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,
RA Hall J., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class II
RT region.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-1678 FROM N.A. (ISOFORM 7).
RC STRAIN=FVB/N. and 129/SV; TISSUE=cartilage;
RX MEDLINE=97135795; PubMed=8981332;
RA Vandenberg P., Vuorio M.M., Ala-Kokko L., Prockop D.J.;
RT "The mouse colla2 gene. Some transcripts from the adjacent rxr-beta
RL gene extend into the colla2 gene.";
RN [3]
RP SEQUENCE OF 1-624 FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).
RC STRAIN=129/SV;
RX MEDLINE=95138212; PubMed=7836472;
RA Tsunaki N., Kimura T.;
RT "Differential expression of an acidic domain in the amino-terminal
RT propeptide of mouse pro-alpha2(XI) collagen by complex alternative
RL splicing.";
RN [4]
RP J. Biol. Chem. 270:2372-2378(1995).
RP SEQUENCE OF 1-8 FROM N.A.
RC STRAIN=129/SV; TISSUE=Liver;
RX MEDLINE=96427460; PubMed=8830784;
RA Tsunaki N., Kimura T., Mateui Y., Ochi T.;
RT "Separable cis-regulatory elements that contribute to tissue- and
RT site-specific alpha 2(XI) collagen gene expression in the embryonic
RL mouse cartilage.";
RN [5]
RP J. Cell Biol. 134:1573-1582(1996).
CC -!- FUNCTION: May play an important role in fibrillogenesis by
CC controlling lateral growth of collagen II fibrils (By similarity).
CC -!- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),
CC alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational
CC modification of alpha 1(XI). Alpha 1(XI) can also be found instead
CC of alpha 3(XI)=1(I) (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Comment=Additional isoforms seem to exist;
CC Name=1; Synonyms=E56789;
CC IsoId=Q64739-2; Sequence=Displayed;
CC Name=2; Synonyms=E5689;
CC IsoId=Q64739-3; Sequence=VSP_007346;
CC Name=3; Synonyms=E5789;
CC IsoId=Q64739-4; Sequence=VSP_007345;
CC Name=4; Synonyms=E569;
CC IsoId=Q64739-5; Sequence=VSP_007346; VSP_007347;
CC Name=5; Synonyms=E589;
CC IsoId=Q64739-6; Sequence=VSP_007345; VSP_007346;
CC Name=6; Synonyms=E59;
CC IsoId=Q64739-7; Sequence=VSP_007345; VSP_007346; VSP_007347;
CC Name=7;

CC CC IsoId=Q64739-1; Sequence=VSP_007345, VSP_007347;
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF100956; AAC69905.1; -
CC EMBL; U16789; AAA67751.1; -
CC EMBL; U16790; AAA67752.1; -
CC EMBL; D38412; BAA18910.1; -
CC EMBL; D84066; BAA12208.1; -
CC MGD; MGI:98447; Coll1a2.
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR008985; Cona_like lec gl.
CC InterPro; IPR008885; Fib_collagen_C.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR003129; TSPN.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 18.
CC Pfam; PF02210; TSPN; 1.
CC ProDom; PD000007; Clg_helix; 1.
CC ProDom; PD002078; Fib_collagen_C; 1.
CC SMART; SM00038; COLFI; 1.
CC SMART; SM00282; LamG; 1.
CC SMART; SM00210; TSPN; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Alternative splicing; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1500
FT PROPEP 1501 1736
FT DOMAIN 31 214
FT DOMAIN 215 486
FT DOMAIN 487 1500
FT DOMAIN 1501 1736
FT CARBOHYD 1604 1604
FT VARSPLIC 267 292
FT
FT VARSPLIC 293 313
FT
FT VARSPLIC 314 373
FT
FT CONFLICT 536 536
FT CONFLICT 621 621
FT CONFLICT 704 705
FT CONFLICT 797 797
FT CONFLICT 843 845
FT CONFLICT 854 854
FT CONFLICT 876 876
FT CONFLICT 889 889
FT CONFLICT 922 922
FT CONFLICT 1005 1005
FT CONFLICT 1253 1253
FT CONFLICT 1386 1386
FT CONFLICT 1522 1522
SQ SEQUENCE 1736 AA; 171535 MW; 18D792D4A3387C61 CRC64;
Query Match 2.4%; Score 319; DB 1; Length 1736;
Best Local Similarity 20.8%; Pred. No. 0.0015;
Matches 405; Conservative 153; Mismatches 739; Indels 652; Gaps 101;
Qy 600 QSABLAMELNESRRWTEEM-----ETAKKGLLEHGRNWSAIARMVGSKTVSQCKN 651
```

Db 107 QGVQOGLGLGRPVRFLEYDQGRPOASQPIFRGLSLADCKWHVAVAKGQSVT- 162
Qy 652 FYFNKKR-----ONLDEILQOHKLAKWERNARRKKKAPAAEAEAAFPVVEDEM 705
Db 163 LIVDCKRVTRPLRSHVPLDTHGVVIFGAH-----ILDEVF 201
Qy 706 EASGVSGNEBEM--VEBAEALHASGN----EVRPGCSGPATVNNSSDTSIPSPHTEA 759
Db 202 E-----GDVQELLVWPGVQAAVSCQKOLECEREQDGPOT-----QKPH- 242
Qy 760 KDTQNGCP-KPPATLGDGPPGPPRRTSRAPTEPTASATGAPTTPPAPSPSAP 818
Db 243 --RAQSPKKEPARLHKP-QSQEPQKOPTESLYDYEPYDVMTTGTAPDYQYPTFG-- 297
Qy 819 PVVPKKEKEEETAAPVVEGEE--QKPPAAEELAVD--TCKAEPEV-----XS 864
Db 298 -----EEEGVLESPLPFLBEEQTDLOVSTADSFQAEYEGEGTDSFAGFYDYTYGYGD 352
Qy 865 ECTEEAEGBAKGDAAEATAEG--ALKAEK-----KEGSGRATTAKSSGA 911
Db 353 DYREETELGALSAETAHSAVAHPRGLKGEKEPAVLEPGMVEGPPGPEGPAGLAGP 412
Qy 912 PQDSDSSATCSADEVDEAGCDKNRLLSRPSLLTFTGDP-----RANASPOKPL 961
Db 413 PGIQGNPGPVG----DFGERGPPGR--AGLPGSDGPPGPGTSLMLPFRFGSGGDKGPV 466
Qy 962 DLKQLKQRAAAIPIPIQVTKVHPPPREDAAATKAPAPPPQNLOPESD--APQPGSGSP 1019
Db 467 VAAQEAQAAILQQAALA-LRGPPGPMGYTGPPGLQPGSGPLKGESGDLGPQ----GP 521
Qy 1020 RGRSRSPAPADK-----EAPAAEAQKLPDPPC-----WTSGLP-----F 1055
Db 522 RGPQGLTGPPFKAGRGRAGADGARGMPGPMKGDGDFGLPGLPGEKGQRGDTGAQGL 581
Qy 1056 PVPREVIRKASHPADPESAFVAPGCHPLPLGLHDTARPVLPR-PPTISNPPPLISSAXH 1114
Db 582 PGPPGE--DGERGDDGETRGLFOESGFRGL-----LGPKGPPGPGPPGV----- 626
Qy 1115 PSVLERQIGAIQSGMSVOLHVPYSBHAKAPVGVFTMGLPLPMDPKLAPFGVSGVQEOQLSP 1174
Db 627 -----RGMD-----GPHGP-----KGS LGP 641
Qy 1175 RGOAGPPESLGVPTAQEASVLRGALG-----SVPGSITKGIPI 1214
Db 642 QSEGPPOGQTPGAQGLPGPOG-AIGPHGEKGARKGKPLGPMFGSDGLPGHPKEGPPG 700
Qy 1215 TR---VPSDSAITYRGSITHGTPADVLYKGTITRIIGEDSPSLDRGDSLP--KGHVI 1269
Db 701 TKGNQGPSGP---QGPLGYGPRGVKGVGIRGLKGHKG---EKG-EDGFPFGKGDIG 751
Qy 1270 YEGKKGHLVYEGGMSVTQCKSEDG-----RSSGPPHETAAPKRTYDME----- 1315
Db 752 VKGDRGEV-----GVPSRGEDGPEGKGRGTGP---TGDPGPTGLMGEKGLGVGLP 801
Qy 1316 ---GRVG-----RAISSAIEGLMGRAIPPERHSPHLLKEQHHTRGS- 1355
Db 802 GYFGQGGKSLGFPFGFAGSEKAGRLSGKSGRGERGTGPRQGRGPRGATGKSGAK 861
Qy 1356 -----TGQIPRSYVEAQEDYLREAKLLKREGTP-----PPPPPSRLDTEAYK 1398
Db 862 GTSGDGPHGPPGEGRLPCPQ-----GPNGFPKPGKPPGAGKDGKLPCHP 906
Qy 1399 TQ-----ALGPLKLPKPAHEGLVATVKEAGRSIHIPRELRHT-PELPLAPRPLKEG 1449
Db 907 QOREVGVGQTKGP-----PGPPGVVPGQGTAGES---GPMGERHSGFPGPPGQGLPGT 959
Qy 1450 SITQGTPLKYDTGASTGSKKHVRSILIGSP-----RTEPPVHLDVWADARALERA 1502
Db 960 SKEGT--KGDGPP-----GAFKDGKAGLGRGFPGERGLPTAGGPKLG- 1003
Qy 1503 CYESLSKRPCTASSG--GSIARGAPVIVPELKGKPRQSPLTVEDHGAPPAGHLPRGSPV 1560

Db 1004 --NEGPAGPPGAPSGPGRGAAGSGGPTGPP--GRPG-----PQGP 1042
Qy 1561 TWREPTPLQOEGSLSSSKASQDRKLTSTPRBIAKSPHSTVPEHHPHPIPSVYHLLRGVSG 1620
Db 1043 AAGEKGVPGKXGPIG-----PTGRDGVQGPVCL-----PGAPG- 1081
Qy 1621 VDLRYSHPTPLAFDPTSPRGIPLDAAAAYLPRHLANPTYPHLYPYLYRGYDPTAALE 1680
Db 1082 EDGDKGEV-----GDPGQKGTGKNKGEGHPGPPGP-IGPV-----GQGAAG 1125
Qy 1681 NRQTIINDYITSQQMHNTATAMARADMLRGLSPRESSLALNVAAGPRGIIDLSQVPHL 1740
Db 1126 GEPG-----ARGPQGHFG-----AKGDETRGFNGPPGPIGLQGLPSPG--EKGETGDG 1173
Qy 1741 PVLVPPPTGTATAMDRLAYLPTAPQPFSSRHSSPLSPG-----PTHLTPTTTSSS 1794
Db 1174 GPMGPPGPPGPRG-----PAGPNGADGPOG-----SPGVGNLGPGEKGEFEGSGSP 1221
Qy 1795 ERERDRDRDRDRERREKSIILTSITTVYEHAPIWRPGTEQSSGSGSGSGGG--GSSSRPA 1852
Db 1222 GVQGEVGVKGRGERGEGK--ESQAGEAGP---PGPKGPTGDNKPKGNPVPVFGDPG 1276
Qy 1853 SHSHAHQSPISPRTDALQORPSVLHNTGMKGIITAVEPSKPTVLKSTSTSSVPVPAAT 1912
Db 1277 PGGA-----GPRQDGAK-----GDRGEDG-----EPGQPGSGPTGENPPGLGK 1319
Qy 1913 FPPA-THCPIGLGTLGVYPTLMBPVLLPKAPVARPERPRADYTHAFIAKPPARSGL- 1970
Db 1320 RGPAGTGPPEG--RQGEKGAKGDPGAV--GAPKGTGP-----VGPAGLAGKPGDGLRG 1369
Qy 1971 -PASSPSKG-----SEPRPLVPP-----VSGH-----ATIARTPAKNLAP 2004
Db 1370 LPSGVQOQGRPGATGQAGPPGVPPOLPGLRDAGAKGKHPGLIGLIGTGEQGEKG 1429
Qy 2005 HHASPPDPAPASADPHREKTSQSPFSIQEELRSLYHSGSSYSEGEVVPVSPVSSPSL 2064
Db 1430 DRGLPGFQGSF-----ELDKSHLSEGLRKPQGPVKLGEEAAHLPHLRPLPESOP- 1463
Qy 2065 THDKGLPKHLE-----ELDKSHLSEGLRKPQGPVKLGEEAAHLPHLRPLPESOP- 2114
Db 1464 PGFSG-PKANGATGAPKGEKVQGP--PGHPGP--PGEV-----IQPLPIQMPKKT 1512
Qy 2115 ---SSPLIQT-----APGVKHQRVWTLAQHISVI-TQDYTRHHPPQOLSAPLA 2161
Db 1513 RRSVDGSKLIQDEEAVPTGAGPSA-----GLEHIFGSLDSLRREEIQMRR--PA 1561
Qy 2162 PLYSPFGASCPLDLRRP--PSDLYLPPPDHG----- 2191
Db 1562 GTQDSPARTQDILKCHPELPDGEYVWDPNQCARDAFVFCNFTAGGETCVTPRDDVTQ 1621
Qy 2192 ---APARGSP-----HSE-----GKRSPEPNKTSVLGGGEGDIEPVSP 2227
Db 1622 FSVVDSEGSFVGVVQLTFLRLLSVAHQDVSVPCSVSODGPK--LRGANEDELSPETS 1679
Qy 2228 PGMWTEGSHRSADVPLLYRDGEQTEPSR 2256
Db 1680 P-----YVKEFRDCQCTQQR 1695

RESULT 91

GPI_CHLRE

ID GPI_CHLRE

STANDARD; PRT; 555 AA.

AC Q9FFQ6; Q03927;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich glycoprotein 1).

GN Gpi.

OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

OC Chlamydomonadaceae; Chlamydomonas.

OX NCBI_TaxID=3055;

101 SEQUENCE FROM N.A.
102 MEDLINE=21159092; PubMed=11258910;
103 Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
104 Goodenough U.W.;
105 "Glycosylated polyproline II rods-with-kinks as a structural motif in
106 plant hydroxyproline-rich glycoproteins";
107 Biochemistry 40:2978-2987 (2001).
108 [2]
109 PARTIAL PRELIMINARY SEQUENCE FROM N.A.
110 MEDLINE=91017504; PubMed=1699225;
111 Adair W.S., Apt K.E.;
112 "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
113 encoding cell wall hydroxyproline-rich glycoproteins";
114 Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359 (1990).
115 -1- FUNCTION: Major component of the outer cell wall w6 (crystalline)
116 layer.
117 -1- SUBUNIT: Associates with GP2 and GP3.
118 -1- PTM: N-glycosylated and O-glycosylated.
119 -----
120 This SWISS-PROT entry is copyright. It is produced through a collaboration
121 between the Swiss Institute of Bioinformatics and the EMBL outstation -
122 the European Bioinformatics Institute. There are no restrictions on its
123 use by non-profit institutions as long as its content is in no way
124 modified and this statement is not removed. Usage by and for commercial
125 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
126 or send an email to license@isb-sib.ch).
127 -----
128 EMBL; AF309494; AAC45420.1; -
129 EMBL; M58496; AAC69706.1; ALT_SEQ.
130 GlycoSuiteDB; Q9FP06; -
131 InterPro; IPR003882; Pistil_extensin.
132 PRINTS; PR01218; PSTLEXTENSIN.
133 Glycoprotein; Repeat; Signal.
134 SIGNAL 1 29 POTENTIAL.
135 CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GP1.
136 DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSXP REPEATS.
137 DOMAIN 259 279 POLY-PRO.
138 CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
139 CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
140 CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
141 SEQUENCE 555 AA; 54219 MW; 64584490465502F5 CRC64;
142
143 Query Match 2.4%; Score 317; DB 1; Length 555;
144 Best Local Similarity 24.4%; Pred. No. 0.00059;
145 Matches 149; Conservative 39; Mismatches 277; Indels 146; Gaps 21;
146
147 717 MVTEAEALHASGNEVPRGECGSPATVNSDTEIPSPHTAEAKDTQNGKPKPATIGAD 776
148 18 MVVLAFAVASANAQCVPGGIFNCPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 77
149 777 GPPGPPPTP---PRRTSRAPETPTASEATGATPP-PAPPPSPAP-----PPVVPKE 825
150 78 PPSGAPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 137
151 826 EKEBETAAAPVBERGEEQKPPAAELAVDTGKAEPEVKSECTEAEAGKAGKDAEAAEA 885
152 138 SPSPAPPPLPPSPAPPSPSPVPPV-----SPSPVPPSPAPPSPPTPPSPSPVPPSPA 190
153 886 TAEGALKAEKGGSGRATTAKSSGACQDSSTSCADEVDEAGGDKRLLSPRSL 945
154 191 PPSAPPVPPSPAPPSPAPPVPPSPAPPSPAPPSPAPPSP-----PSPA 232
155 946 TPTGDRPRANASQKPLDLKQLKQRAAIPPIQVTKVHPREDAAPTKAPPAPPPQNL 1005
156 233 PPSGPPAPPSPVPP-----SPAPSPAPPSPKPP---APPPPPPPPPPPPPPP 279
157 1006 QPESDAPQQRSGSRGKSRSPAPPADKEAFAAAEQKLPGLPPFPVPPREV-IK 1064
158 280 FP-ANTMPP--SPSPSPSPAPT-----PPTPP--SPSPSPSPVPPSPAPVP 322
159 1065 ASHAPDPSAFSAPPGHPLGLLHDHTARP-VLPRPTTINPPPLISSAKHPSVLRQIG 1123
160
161

Db 323 PSPAPPSPAP---SPPPSPAP-----PTPSPSPSPSPSPSPSPSPSPSPSPSPPIP----- 371
Qy 1124 AISQMSVQLHVPYSEHAHAPVGPVTMGLPLPMPKPLAPSPGVKQEQLSRPGAGPPES 1183
Db 372 -----SPSPKSPSPSPVAVKLVWADDAIAFDLLNGTSTRPGSASRWGEPDI 417
Qy 1184 LGV-----PTAQEASVLRGTALGSVPGSI----- 1208
Db 418 AGTKCKGNLKGWMPKPSRNPWGQAVFSGGRTVGSVANVTIRVAFATEKPKALIYSSIELV 477
Qy 1209 --TKGIPSTRVP-----SDSAITYRGSITHTGTPADVLKGTIT-----RIIGEDSPSR 1254
Db 478 VYNTGATLIRPIAANVTIRSOIRCPGFLTYGTTPIAGYPTCIDATTNPWNKIAG-----VR 533
Qy 1255 LDRGREDSLPK 1265
Db 534 INMGAGNKKPK 544
RESULT 92
ID_PSC DROME STANDARD; PRT; 1603 AA.
AC P35820;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Posterior sex combs protein.
GN PSC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92018190; PubMed=1833647;
RA Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
RT "Drosophila genes Posterior Sex Combs and Suppressor two of zeste
RT encode proteins with homology to the murine bmi-1 oncogene";
RL Nature 353:351-353 (1991).
CC -1- FUNCTION: The Polycomb group (Pc-G) genes are needed to maintain
CC expression patterns of the homeotic selector genes of the
CC antennapedia (Antp-C) and bithorax (Bx-C) complexes, and hence for
CC the maintenance of segmental determination.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59275; CAA41965.1; -
CC PIR; S17983; S17983.
CC FlyBase; FBgn005624; Psc.
CC InterPro; IPR001841; Znf_ring.
CC Pfam; PF00097; zf-C3HC4; 1.
CC SMART; SM00184; RING; 1.
CC PROSITE; PS00518; ZF_RING_1; 1.
CC PROSITE; PS00089; ZF_RING_2; 1.
CC Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT DOMAIN 47 53 POLY-THR.
FT DOMAIN 83 88 POLY-THR.
FT DOMAIN 91 98 POLY-THR.
FT DOMAIN 145 152 POLY-THR.
FT DOMAIN 184 202 POLY-SER.
FT ZN_FING 265 304 RING-TYPE.
FT DOMAIN 642 651 POLY-SER.
FT DOMAIN 1066 1069 POLY-GLY.
FT DOMAIN 1185 1189 POLY-PRO.

Db 1039 PGAPGPGSPGAGTCKQGRGAGAGQPMGPS-----GPAGARGIAGPQGRGDKE 1092
 Qy 1955 -----TGHAFLEKPPARSGLEPASSKSGSEPPPLVPVPSGHATARTPA 1999
 Db 1093 SGOGERGLKGRGFTGLQGLPGPGPGSGDQSGAGPAGSPGR----- 1135
 Qy 2000 KNLAPHHASPPPPAPPASADPHREKTQSKPFSIQLELRSLGYHGSYSBPGVERPSPV 2059
 Db 1136 -----GPGCPVGPS-----GKGSNGIPGPGPPGR 1162
 Qy 2060 SPSLSLTHDKGLPKHLEELDLSHLEGLRKPQGPVKLGGAALHPLRLPLPSQSSSPL 2119
 Db 1163 GRSGETGPVPPG-----SPGPPGPPGPGGIDMSAFAGLGQREKGPDM 1208
 Qy 2120 LQTAGVKGHORVVTLAHLSEVITDTRH--HPQOLSAPLPAPLYSPFGAGCPVLDLR 2177
 Db 1209 QY-----MRADEADSTRQHDVEV---DATLKSINNOIESIRSPDGRSKNPARTCDQLK 1261
 Qy 2178 RP--PSDLYLPPDPHAGARGSP---HSEGGKRSPEPNKTSV 2214
 Db 1262 HPEWKSQDGYWDNQCTLDAMKVCNMTGETCVYVPATV 1303

RESULT 94

CA25_HUMAN
 ID CA25_HUMAN STANDARD; PRT; 1496 AA.
 AC P05937;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 2(V) chain precursor.
 GN COL5A2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE OF 1-463 FROM N.A.
 RX MEDLINE=89123368; PubMed=2914927;
 RA Woodbury D., Benson-Chanda V., Ramirez F.;
 RT "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
 RT to the structural criteria of a fibrillar procollagen molecule.";
 RJ J. Biol. Chem. 264:2735-2738(1989).
 RN [2]
 RP SEQUENCE OF 398-1496 FROM N.A.
 RX MEDLINE=87146331; PubMed=3029669;
 RA Weil D., Bernard M.P., Gargano S., Ramirez F.;
 RT "The pro alpha 2(V) collagen gene is evolutionarily related to the
 RT major fibrillar-forming collagens";
 RL Nucleic Acids Res. 15:181-198(1987).
 RN [3]
 RP SEQUENCE OF 1227-1496 FROM N.A.
 RX MEDLINE=85289337; PubMed=2411731;
 RA Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
 RT "Complete primary structure of the human alpha 2 type V procollagen
 RT COOH-terminal propeptide";
 RJ J. Biol. Chem. 260:11216-11222(1985).
 RN [4]
 RP SEQUENCE OF 1449-1496 FROM N.A.
 RX MEDLINE=89138450; PubMed=3224983;
 RA Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
 RA Ramirez F.;
 RT "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
 RT located on the long arm of human chromosome 2";
 RL Genomics 3:275-277(1988).
 RN [5]
 RP SEQUENCE OF 208-227.
 RC TISSUE=Placenta;
 RX MEDLINE=92239022; PubMed=1571108;
 RA Mann K.;
 RT "Isolation of the alpha 3-chain of human type V collagen and
 RT characterization by partial sequencing.";
 RJ Biol. Chem. Hoppe-Seyler 373:69-75(1992).

[6]
 RN SEQUENCE OF 288-297 AND 606-617.
 RP TISSUE=Bone;
 RX MEDLINE=94237164; PubMed=8181482;
 RA Moradi-Ameli M., Rousseau J.C., Kieran J.P., Champlaud M.F.,
 RA Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
 RT "Diversity in the processing events at the N-terminus of type-V
 RT collagen.";
 RL Eur. J. Biochem. 221:987-995(1994).
 RN [7]
 RP DISEASE
 RX MEDLINE=98087576; PubMed=9425231;
 RA Michalickova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;
 RT "Mutations of the alpha2(V) chain of type V collagen impair matrix
 RT assembly and produce Ehlers-Danlos syndrome type I.";
 RL Hum. Mol. Genet. 7:249-255(1998).
 RN [8]
 RP VARIANT EDS-II ARG-960.
 RX MEDLINE=98455031; PubMed=9783710;
 RA Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,
 RA Burrows N.P.;
 RT "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type
 RT II.";
 RL J. Med. Genet. 35:846-848(1998).
 CC -1- FUNCTION: Type V collagen is a member of group I collagen
 CC (fibrillar forming collagen). It is a minor connective tissue
 CC component of nearly ubiquitous distribution. Type V collagen binds
 CC to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
 CC -1- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
 CC most tissues and trimers of one alpha 1(V), one alpha 2(V), and
 CC one alpha 3(V) chains in placenta.
 CC -1- PPM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
 CC type I (EDS-I) [MIM:130000]; also known as Ehlers-Danlos syndrome
 CC gravis. EDS-I is an autosomal dominant connective-tissue disorder
 CC characterized by loose-jointedness and fragile, velvety,
 CC stretchable, bruisable skin that heals with peculiar 'cigarette-
 CC paper' scars. Inheritance is autosomal dominant.
 CC -1- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
 CC type II (EDS-II) [MIM:130010]; also known as Ehlers-Danlos
 CC syndrome mitis. Inheritance is autosomal dominant.
 CC -1- SIMILARITY: Contains 1 VWFC domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J04478; AA51859.1; -
 DR EMBL; X04758; CAA28454.1; -
 DR EMBL; M11718; AAA52058.1; -
 DR FIR; A31427; GCHU2V.
 DR PDB; 1A9A; 18-NOV-98.
 DR Genew; HGNC:2210; COL5A2.
 DR MIM; 120190; -
 DR MIM; 130000; -
 DR MIM; 130010; -
 DR GO; GO:0005588; C:collagen type V; TAS.
 DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; vwc; 1.
 DR ProDom; PD000007; C1g_helix; 5; 1.
 DR ProDom; PD002078; Fib_collagen_C; 1.

DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; WVC; 1.
DR PROSITE; PS01208; WVC; 1; 1.
DR PROSITE; PS0184; WVC; 2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Ehlers-danlos syndrome;
KW Disease mutation; 3b-structure.
FT SIGNAL 1 26
FT CHAIN 27 1226
FT PROPEP 1227 1496
FT DOMAIN 39 97
FT MOD_RES 290 290
FT MOD_RES 293 293
FT MOD_RES 296 296
FT MOD_RES 608 608
FT MOD_RES 614 614
FT MOD_RES 960 960
FT VARIANT
FT CONFLICT 292 292
FT CONFLICT 1418 1418
FT CONFLICT 1438 1438
FT CONFLICT 1460 1460
FT CONFLICT 1496 1496
FT SEQUENCE 1496 AA; 82827C17A8644F5A CRC64;

Query Match 2.4%; Score 315.5; DB 1; Length 1496;
Best Local Similarity 21.2%; Pred. No. 0.0016;
Matches 354; Conservative 112; Mismatches 543; Indels 663; Gaps 91;

Qy 728 GNEVPRGECSPATVNNSSDTESTPSHTEAAKQTQNGKPPATLADGPP-----PGP 782
Db 138 GSQGRGGE-RGP-----KRGPRGPQ--GIDGPGVPGQFGA 172
Qy 783 PTPPRRTSR-----APIEP-----TPASEAT 803
Db 173 PGPGHSHPGDGLSRPFSQAAGLDEKSLGQVGLMPCSVGPRGPGQLGQGG 232
Qy 804 GAPTTPAPPSPAPPPVVPKKEEETAAAPPVEGEQKP-----PAABELAVDTGKA 858
Db 233 AGPTGPPG--EPGDGPGMPGPIGSRGPE--GPPGKPGEDGPRGNPGVEGVFAGSPGAR 287
Qy 859 EEP-----VKSECTEAEKGP-----AKGDAAEATAEAGALKAEKKEGGSGRATTA 906
Db 288 GFPGAGPLGLGHRHKGLEGPGEVGAAGSGEAGTGMGAMGPLGPRGMFGERGL 347
Qy 907 KSSGAPQDSDSATCSADEVDEAEGDKNRLLSRPSLLTPTGDPANASQKPLDLKQL 966
Db 348 GPQAGPGORG-----AHG-----MPGKPGMGLGIPGSSGFFGNP-----383
Qy 967 KQRAAAIPPIQVTKVHEPPREDAAPTAPAPPQQNLOPESDAPQPGSSPRGKSRSP 1026
Db 384 -----GMKGEAGTGAEGPGGQ-----RGETGPP 410
Qy 1027 APPADKEAFABAOKLPGDP--PCWTSGLPFPVPPREVTKASPHADPSAFYAPPGHPLP 1085
Db 411 GPVGS-----PGLPGAIGTGTGPPGKGP-----TGSPGTSGPPG--SAGPPGSPGP 454
Qy 1086 LG-LHDTARPVLPRTPTI-----SNPPPLISSAKHPSVLERQIGALSQMSVQLHVPY 1137
Db 455 QGSTGPGNSGLPGDGPFGKGAEPKGP-----HGIGPIGPPG 495
Qy 1138 SEHAKAPVG--PVTMGLPLPMDPKKLAPSGVKQBLSPRGQAGPPESLGVPVTAQEAVALR 1196
Db 496 BEGKRGPRGDPGLGPPGPGV-----ERGA--GNRGFPDGLPGPKGAQGER 542
Qy 1197 GTALGSPVGGSIYKIGISTRVPSDSALTYRGSITHTGTPADVLVYKGTITRI--IGEDSPSR 1254
Db 543 GPVSSGPKGS--QCDPR--PGBPGLPGARGLT--GNPGVQGPREGKLGPLGAPGED-----593
Qy 1255 LDRGREDSLPKGHVYEGKKGHVLSYEGGMSVTCQSKEDGRSSGSPPHETAAPKRTYDMM 1314
Db 594 ---GRPG--PPGSIIGKGPQGTM-----GLPGKSGNSGND-----PGKPGAGNPG-----V 634

RESULT 95

UN89_CABEL

ID UN89_CABEL

AC 001761; Q17362;

STANDARD;

PRT; 6632 AA.

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT assembly, encodes a giant modular protein composed of Ig and signal
RT transduction domains.";
RL J. Cell Biol. 132:835-848(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U33058; AAB00542.1; -;
DR EMBL; AF003131; AAB54132.2; -;
DR PDB; 1FHO; 20-DEC-00.
DR WormPep; C09D1.1; CE30426.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001849; PH.
DR InterPro; IPR007850; RCSD.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 47.
DR Pfam; PF00169; fh; 1.
DR Pfam; PF05177; RCSD; 5.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGC2; 23.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PSS0010; DH_2; 1.
DR PROSITE; PSS0835; IG LIKE; 49.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
DR PROSITE; PSS0002; SH3; 1.
DR Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4583 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.

RESULT 96
MUC1 HUMAN
ID MUC1 HUMAN STANDARD; PRT: 1255 AA.
AC P15941; P15942; P17626; Q14128; Q14876; Q16437; Q16442;
AC Q16615; Q9BXA4; Q9UE75; Q9UE76; Q9UQL1; Q9Y4U2;
DT 01-JAN-1990 (Rel. 13, Created)
DT 10-OCT-2003 (Rel. 14, Last sequence update)
DE Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)
DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE (Tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut-
DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DE DF3) (CD227 antigen).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pancreas;
RX MEDLINE=90368716; PubMed=2394722;
RA Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=90202794; PubMed=2318825;
RA Lightenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Feat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of the human polymorphic epithelial mucin
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Feat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 5).
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,
RA Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transcribed gene and over-expression in breast cancer
RT tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=91033045; PubMed=1688329;
RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RN [8]
RP SEQUENCE FROM N.A. (ISOFORM 7).
RX MEDLINE=95010060; PubMed=7925397;
RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
RA Keydar I., Hilkens J., Wreschner D.H.;
RT "Characterization and molecular cloning of a novel MUC1 protein,
RT devoid of tandem repeats, expressed in human breast cancer tissue.";
RL Eur. J. Biochem. 224:787-795(1994).
RN [9]
RP SEQUENCE FROM N.A. (ISOFORMS 6; 7 AND 8).
RX MEDLINE=97355747; PubMed=9212228;
RA Oosterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,
RA Finstad C.L.;
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
RT cancer cell lines and demonstration of a new short variant form
RT (MUC-1/Z).";
RL Int. J. Cancer 72:87-94(1997).
RN [10]
RP SEQUENCE FROM N.A. (ISOFORM 7).
RX Zhang L.X., Li C.H.;
RT "Molecular cloning of an isoform of MUC1, MUC1/Y.".
RN Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A. (ISOFORM 9).
RC TISSUE=Epithelial cancer;
RX Zhang L.X., Li C.H., Sun L.Y., Yue W.;
RT "Cloning of a new potential secreted short variant form of MUC1 mucin
RT in epithelial cancer cell line.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [12]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [13]
RP SEQUENCE OF 1-160 FROM N.A. (ISOFORM 2).
RX MEDLINE=90088473; PubMed=2597151;
RA Abe M., Siddiqui J., Kufe D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [14]
RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM 2).
RC TISSUE=Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";
RL Int. J. Cancer 66:55-59(1996).
RN [15]
RP SEQUENCE OF 1-89 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA Lee L.N., Luh K.T., Wu C.W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues.";
RL Oncology 53:118-126(1996).
RN [16]
RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=Breast carcinoma;
RX Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
RN [17]
RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.

DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; Clg_helix; 7.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Signal.
FT SIGNAL 1 25
FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.
FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.
FT DOMAIN 1485 1712 NON-HELICAL REGION (NC1).
FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).
FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).
FT DISULFID 1549 1555 BY SIMILARITY.
FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).
FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).
FT DISULFID 1658 1665 BY SIMILARITY.
FT CARBOHYD 138 138 N-LINKED (GLCNAC...).
FT CONFLICT 471 471 R -> P (IN REF. 2).
FT CONFLICT 683 683 A -> G (IN REF. 2).
FT CONFLICT 1575 1575 M -> I (IN REF. 5).
FT CONFLICT 1663 1663 G -> H (IN REF. 9).
FT CONFLICT 1701 1701 H -> G (IN REF. 9).
SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;

Query Match 2.4%; Score 312.5; DB 1; Length 1712;
Best Local Similarity 19.8%; Pred. No. 0.0023;
Matches 407; Conservative 146; Mismatches 704; Indels 803; Gaps 103;

Qy 709 GVSNGEEMVEEABALHASGNEVREGCSGPATVNNSSDTSIPSPTEAAKDTGQNGPK 768
Db GLQGRKDKGKRG---APGVTGPKDV-GARGVSGPFGADGIFG-----HFGQGGPR 129

Qy 769 -PPATLGADGP-----PPGPTTPRTSRDIEFTPAEATGATPPAPPSPAPPPV 822
Db GRPYDGCNGTQSDSGPQGP-----PGSEGTGPPGPKGKGEPYAL 174

Qy 823 PKKEBETA-----AAPVEGEEOK-----PPAAEELAVDTGKA 858
Db PKERDRIYRGPBGLVFGPPGPRGHVGMGPVAGPRGPPGPPGKQGNRGLG 234

Qy 859 EEPVKSECTEBAERG-----AKGKDAEAAEATAEG-ALKAE 894
Db FYGVKKGVDVGPQNGPISDTLHPITAPGVTFHPDYKKGSGEPCIRGSLKGE 294

Qy 895 KK-----EGSGRATTAAGQSDSDSATCSADEV-----EAGGDKN-----RLLSPR 941
Db EGMGFPGLRGYPGLSGEKGSPGQKSGRGLDGYQGPDPGRPKGEAGDPGPPGLPAYSPH 354

Qy 942 PSLLT-----PTGDPANASPKP-----LDLKLKQRAA----- 971
Db PSLAKARGDPGPFAGAQEGSQGEGPDGLGPPGLSIGDQRRGLBEMGPKGFID 414

Qy 972 -AIPPIQVTKVHEPPREDAAFTKDP---PAPPPP-----QNLQESAPQPGS-SP 1019
Db PGIPAL-----YGGPPGDKRGPPGPPGLPGPDGFLGLKCAKGRAGPGLPGSPGA 470

Qy 1020 RGKRSAPPADKAEFAAEAKQLPGDPPCTWTSGLPFPVPPREVTKASPHADPSAFS--- 1076
Db 471 RG-----PKMGKDAGECRCTEGD-----EAIKGLPLGPKGFAGIN 508

Qy 1077 -----YAPPGH-----PLPLGLHDTAR----- 1093
Db GEPGRKGDGDPGQHLGPFGLKGVNIGAPGPKAKGDSRTITTKRGSGQGVGPV 568

Qy 1094 -----PVLPRPP--TISNPPPLISSAKHPSVLERQIGAI SQMSVOLH 1134
Db 569 GMKGDGSGPDGDLGDFGLPGFPDGIKGGP-----GDPGYFG----- 607

Qy 1135 VPYSEHAKAPVGPVTMGLP-----LPMDDKULAP----- 1163
Db

Db 608 IPGKTGTPGEMGPPGLGLGLKQGRGPGDAGLPGFPGLGPPGAGTGGIDCDTDVGR 667
Qy 1164 -FSGVKQEQLSPRQAGP---PESLGVPTAQEASVLGRTALGSPGSGITKGIESTRVPS 1219
Db 668 AVGDRQEAIQPGCIAGPKGLPGLPGGPGTGAAGLKG-----IPGAGADGGGPR--- 719
Qy 1220 DSAITYRGSITHTGTPADVLYKGTITRIIGEDSPRLDRGRRED-SLPKGHVYIEKGKHVL 1278
Db 720 -----GLPGDA-----GREGFGPPGFIGRSGKGA- 746
Qy 1279 SYEGGMSVTQCKEDGR-----SSSGPPHETAAPKRTYDMMEGRVRGAISSASIEGLM 1331
Db 747 -----GL-----PGDGSPPGILPGDPPGGERGLGCEVLGAQPGPRG-----DAGVPQGP 793
Qy 1332 G-RAIPPERHSPHLKEQHIRGSIITQIGIPRSYVEAQEDYLRRKALKRGTGTPP--- 1386
Db 794 GLKGLPGDRGPP-----GFRGS--QOMPOM-----PGLKGQGLPGSPQGP 832
Qy 1387 -----PPPSRDLTEAYKTOALGPKLK--PAHEGLVATVKEAGRSIHIREELRHTPELP 1440
Db 833 GLYGPGLHGFPGAPQGE--GPLGLPGIPGREGL-----PGD--RGDPGDT 874
Qy 1441 LARPL-KEGSI TCGTPLKYDTG--ASTGSKKH-----DVRSLIGSPG-----RTFPV 1487
Db 875 GAGPVGKMG-----LSGDRGDAGFTGQGHFPGSPGFKGIDGMPGTGPKLGRGSPGM 927
Qy 1488 HPLDVMADARALERACRYEESLKSRRPGTASSGSGSIARGAPVIVPELKGPRQSPLTYEDHG 1547
Db 928 DGFQGM-----PGLKGRPPGSGKEAGPFGIPLKCLAGEP----- 964
Qy 1548 APFAGHLPRGSPVTMREPTPL-----OEGSLSSSKASQDRKLTSTPREIAKSP 1596
Db 965 -GFKG--SRGDGPPGPPPVILPGMKDIKGEKDEGPMGLKGYLCAKGIQMP----- 1014
Qy 1597 HSTVPEHHPHPISPYEHLLRGVSCVDLYRSHIPLAFDPTSPRGIPLDAAAAYLPRHLA 1656
Db 1015 --GIPG-----LSGIPGLGPRGPHIKGVKGDIGVP-GIP----- 1045
Qy 1657 PNPTYPHYLYPPIRGYPDTAALENROTIIINDYITTSQMHNTATAMAQADMLRGLSPR 1716
Db 1046 GLPGFPVAGPPGILTFFG-----FTGSRGD--KGAPGR 1077
Qy 1717 ESSIALNAAAPRGIIDLSQVPHLVLVPPPTGTPATAMDR-LAYLPTAPQPFSSRHSS 1775
Db 1078 AGLYGEIGATGDFG--DIGDTINL-----PGRPGLKGERGTGIPGLKGFCEKGTGEG 1128
Qy 1776 PLS-PCGPTHLYKPTTSSSERERDRDRDREREKSIILTSTTTVEHAPIWR-----P 1829
Db 1129 DIGFPG-----ITGVTGVQPPGLKGQTGFP 1154
Qy 1830 GTEQSSGSSGSG-----GGGSSSRPASHAHQHSPISPTQDALQQRPSVLHNTGMK 1884
Db 1155 GLTGPPGSGELGRIGLPGKGGDDWPGA-----PGLGPPGLR 1193
Qy 1885 GII-----TAVEPSKPTVLRSTSTSSPVRAATFPFATHCPGLGTLG-DVYPTLMEPV 1936
Db 1194 GIRGLHGLPGTKGFPGSP-----GSDIHGDPGFP-----PGRGDPCEANTLPGPV 1241
Qy 1937 LLPKEAPRVARPERPRADTGHAFKAPKPARSGLEPASSPKSEPRPLVPVSVCHAIAR 1996
Db 1242 GVFGQ-----KGDQAGPGERGPPGSPGLQG-----PFGITPPSNISG 1278
Qy 1997 TPAKNLAPH---HASPPDPAPPASADPHREKTSQKPFSTIQEELRSLGVHSGSSYSPG 2052
Db 1279 APDGKAPGIFGLKGYRGPPGPPGSAALPGSKGDTGNP-----GAGGTPGTGKW 1327
Qy 2053 VEPVSPVSSPSLTHDKGLPKHLEELDKSHLEGLRPKQGPVFKLGGAALPHLRPL--- 2109
Db 1328 AGDSGPGRGPGV---GLP-----GEKGR-----GEGGFMGNTGPTGAV 1364
Qy 2110 -----PESQPSSSLLQATAPGVKGHQVRVTLAQHISEVITQDTRHHQQQLSAPLAPLYS 2165
Db 1365 GDRGPKG-PKGDGPGFPGAPGTGAPGIAGIPQKIA---IQPGT-VGPQGRGPPGA--- 1415

QY 2366 NAFNPLNA---SASLPAAMPITRADGRSDHTLSPGGG-----GKAKVSG----- 2407
 DB 2402 STFVPIQAGPQMLTIPAVSVIHRVTGSGDTITTEASGSPNRPTGVAEELSSWPCIPIGQI 2461
 QY 2408 -RPSRRKAKGPA-----PGLA-----S 2423
 DB 2462 HVPGLQNLSPFALQSLTSLGMEITVNLVGLANATVPGQHPHGLNALVAGLOVLNAPAGS 2521
 QY 2424 GDRPPSVSVHSGDCNRRPTPLNRWVEDRPPSAGSTPPPY-----NPLIMRLQ--AGV-M 2476
 DB 2522 SPAPPA-----HIQGLQLNTALPTLTPSVGPAVGTGTPTTAPNSKAMELQMPAGQGH 2577
 QY 2477 ASPPPPGLPAG-----SGPLAGPHHAWDEE 2501
 DB 2578 SAEPQSGSPGQETQTQVSGPSA--DHARPED 2608

RESULT 99
 PGCV MOUSE
 ID PGCV MOUSE STANDARD; PRT; 3358 AA.
 AC Q62059; Q62058; Q9CUUO;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Versican core protein precursor (large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
 GN CSFG2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS V0; V1 AND V2).
 RC STRAIN=C57BL/6; and Swiss Webster; TISSUE=Brain;
 RX MEDLINE=95122551; PubMed=7822336;
 RA Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;
 RT "Multiple forms of mouse PG-M, a large chondroitin sulfate
 RT proteoglycan generated by alternative splicing.";
 RL J. Biol. Chem. 270:958-965(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM V3).
 RC STRAIN=C57BL/6;
 RX MEDLINE=95181355; PubMed=7876137;
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
 RT without a chondroitin sulfate attachment in region in mouse and human
 RT tissues.";
 RL J. Biol. Chem. 270:3914-3918(1995).
 RN [3]
 RP SEQUENCE OF 1-1692 FROM N.A. (ISOFORM V1).
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RX MEDLINE=23354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Knäuper A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lennard B., Lyons P.A.,
 RA Maglott D.R., Maitais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilmink L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,

RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP INTERACTION WITH FBLN1
 RX MEDLINE=9329059; PubMed=10400671;
 RA Asberg A., Adam S., Kostka G., Timpl R., Heinigaard D.;
 RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and
 RT versican.";
 RL J. Biol. Chem. 274:20444-20449(1999).
 CC -!- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -!- SUBUNIT: Interacts with FBLN1.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=4;
 CC Comment-Additional isoforms seem to exist;
 CC Name=V0;
 CC IsoId=Q62059-1; Sequence=Displayed;
 CC Name=V1;
 CC IsoId=Q62059-2; Sequence=VSP_003087, VSP_003088;
 CC Name=V2;
 CC IsoId=Q62059-3; Sequence=VSP_003089;
 CC Name=V3;
 CC IsoId=Q62059-4; Sequence=VSP_003087, VSP_003090;
 CC -!- TISSUE SPECIFICITY: V2 is found only in brain.
 CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D16263; BAA03796.1; -;
 CC EMBL; D28599; -; NOT_ANNOTATED_CDS.
 CC EMBL; D32040; BAA06802.1; -;
 CC EMBL; AK014525; BAB29411.2; -;
 CC HSSP; P01132; 1EPG.
 CC MGD; MGI:102889; Cspg2.
 CC InterPro; IPR000152; Asx_hydroxyl_S.
 CC InterPro; IPR000742; EGF-2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF-like.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR003599; Ig.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR000436; Link.
 CC InterPro; IPR000538; Sushi_SCR_CCP.
 CC Pfam; PF00008; EGF; 2.
 CC Pfam; PF00047; Ig; 1.
 CC Pfam; PF00059; Lectin_c; 1.
 CC Pfam; PF00084; sushi; 1.
 CC Pfam; PF00193; Xlink; 2.
 CC PRINTS; PR01265; LINKMODULE.
 CC ProDom; PD000918; Link; 2.
 CC SMART; SM00032; CCP; 1.

QY 951 PRANASQKPLDLKQLKORAAAIPIQVTKVHPPREDAAPTAPAPPPQN-----LQ 1006
Db 1389 QLVTVVPKDP-----EAAERGOYEVAPSNFDDSATDTHQFILAE 1432
QY 1007 PESDAPQOGSSPRGKRSRAPPADKFAFAEAQKLPDPPCWTSGLPFFVP----- 1058
Db 1433 TEGSTTQFKSKEG-----TELLEITWKETYPETPDHVSSEDPVFPFTLSSHCK 1484
QY 1059 ---PREVIKASHADPSAFSPCHPLPLGLHDTA-----RPVLP----- 1098
Db 1485 TTRWSEITSSNLENP---VHKQP-KPVLPFEESGEGAIEQASQETILSRATEVAL 1540
QY 1099 -----PPTISNPPPLISSAKHPSVLERO-----ICAISSQGMV-----QLHVPY 1137
Db 1541 GKETOQSPTLST--SSILSSSVSNVLEERPLITLIGISQDSNSTIESWVEITPQTKF 1599
QY 1138 SEHAKAPVGVMTWGLPLPMDPKKLAPFSGVKQE-----QISPRQAGPPESLGVPTAQ 1190
Db 1600 SSSAPI-----IEGSEVENKNKIFNMVTDLPQRPDPTDLS-PLDM 1642
QY 1191 EASVLRGALTALGVPGGSIKTIKIPSTRVPSDSALTIRGSIHGTADVLYKGTITRIIGED 1250
Db 1643 SKIMITNHHI-----YIPATIAPLDSK-----PSPDARPTTVWNSNSTSEWVSDK 1688
QY 1251 SPRLDRGREDLPGKHVIYEGKGHVLSYEGGMSVTQCKEDGRSSSGPPHE---TAAP 1307
Db 1689 S-----FEGRK-----KKENEDERGAVNAHQGEVRAAT 1717
QY 1308 KR-----TYDMMEGVGRCAISSASTIEGLMGRAIPPERHSPLHKEQHIIHGSIQGIPIR 1361
Db 1718 ERSDHLLLPTELESSNVDASSDLATWEGFILTPTTESE-----KEMANSTPVFRETIGV 1772
QY 1362 SYVEAQ-----EDYLRRBAKLKREGTPPPPPSRDLTEAVKQALGPLKLKAHEGLV 1415
Db 1773 ANVEAQPFHSSSHPRVQFELTILSGNPPSL-----FTDLGSGDASTGMELITA 1822
QY 1416 ATVKEAGRSIHIPRE-ELRHTEPLPLAPRLKEGS-----ITQGTPLKYD-TGASTT 1466
Db 1823 -----SLFTLDLESETKVKKELEFSPSPSVSEISSFEPTGLTPSTVLDIETAGVMSQ 1874
QY 1467 GSKKHVRLSLGSPKPTPPVHPLVDMAARALERACYEESIKNRPGTASSGSGSIARGA 1526
Db 1875 TSQKTLISEISKP-----TSQSGVRDLTYGF 1901
QY 1527 PVIVELKPROSLTYEDHGAPFAG---HLPRGSPVTMRBPT-----PRLOEGSL 1575
Db 1902 PM-----GDFSGDFSEYPTVSPTMKETVGMGSDDERVDTQTS 1943
QY 1576 SSKASQDKLSTPREIAKSPHSTVPEHHPHPISPYHLLRGVSG---VDLYRSHI--- 1628
Db 1944 SSIPTSDNIYVPVD--SKGPDSTVAS---TTAFWEVMSAARGSGEQLASRVSSGVP 1998
QY 1629 -PLAFD-----PTSIP---RGIPLDAAAAYLPRHLAPN 1658
Db 1999 LPLAVDIFSGTESPYDFSEFEVAATVEANERPVLPTAAGNTVDLTENGIEVNTMS 2058
QY 1659 PTPHYLPYPLIRGVPDTAALE---NRQTIINDYITQQMHNTATAMAQRAQMLRGLSP 1715
Db 2059 LDFPQTEPESKLGKPEVNLKQBEIGRETIVTKERAAQKQTFESLHSSFAPEQITILETQSL 2118
QY 1716 RESSLALNYAAGPRGIDLSOVPHLPVLVPTTP-----GTPATANDRLAYLPTA 1764
Db 2119 IETEFQIS-----DYSMLTILKYITNKEVEEGMSIAHMSTPGPGIKOLESYTH 2169
QY 1765 POPFSRRHSSPLSPGQPHLTHTKPTTTSSSERDRDRDREREKSLTITTTVEHA 1824
Db 2170 PEAPCKSHSFS-----ATALVETSGAAR-----SVLMDSSSQEEE 2204
QY 1825 PI---WRPG---TFQSSGSSSGGGGSSSRP-----ASHSHA 1857
Db 2205 SIKLFQKGVKLTNKESSNADISFSLGSGGALPPLPTTSVNLTKQIISTLYAETSHMES 2264

QY 1858 HQHSISPRTOALQORPSVLHNTGMKIITAV-----EPSKPTVLRSTSTSSVRPA 1910
Db 2265 LGTSLGDKMED--HERMEDVSSNEVRMLISKIGSIQDSSTEALDITLTSHTGTTEPT--T 2320
QY 1911 ATFFPATHCPGLGTLGDVYPTLMEPVLLPKEAP-----RVARPERPRADTGHAFKAPAR 1966
Db 2321 STLP-----FVKLMDLERSPKQDPGGGRKPKTHRPQTMGSLGINENSSAS 2366
QY 1967 SGLPPASPS-----KGSEPRPLVPVPSGHA-----TIARTP 1998
Db 2367 EAREGATSPTAFLPQYTSVEMTKHFAPSESPSDFNVNSGEGSEVDTLDLVYTSQTQ 2426
QY 1999 AKN-----LAPH---HASPPPPAPPASADPHREKTKOSKPSQELRLSLGHVGSYSP 2050
Db 2427 ASSGDSNKLASHGLEKHPKSVKTEAGATDV--SPTASAMF-----LHSEYKSSLYPT 2478
QY 2051 EGVFVPSVSPSLTHDKGLPKHLEELDKGHLGE-LRPKQPGVKVLGGEAAHLPHLRPL 2109
Db 2479 STLSTEPYKSPSEGIEDGLQ-----DNTQFEGSTLKPSR-----RKT 2516
QY 2110 PES-----QPSSSPLLOTAPGVKHQRVVTLAQHISEVITQDYTRHHPOQ 2154
Db 2517 TESIIIDDKEDSKDLGLTITITESAIVKSLPELTSKNI-----IIDIDHTK----- 2562
QY 2155 LSAPLPAPLYSF-PGASCPVLDLRRPPSCLYLPDPDGAPARGSPHSE--GGKRSPEPNK 2211
Db 2563 ---PVYEYIPGIQ---TDL---DEIKLESHGSEESLQVQEKYEGAVTILSPTE 2607
QY 2212 TSVLGGGEDGIBVSPPEGMTEPHRSRVAVYPLLRYRDEQTE----- 2253
Db 2608 ESFEGSGDALL-----AGYTOAIYNES-VTP---NDGQAEDISFSPATGIPVSSTETE 2657
QY 2254 -----PSRMGSKSPGNTSQP-----PAFFSKLTESAMVSKKQKINK 2293
Db 2658 LHFTFPTASTLHTIPSKLTTASP-EIDKPNIEAISLDDIFESSTLSGQAI-ADQSEVIST 2715
QY 2294 LNTHNNEPEYNISQPCTEIFNNPFAITGTLTYRSQAVQEHASTNM-----GLEAIR 2347
Db 2716 LGHLEKTOEYEBEKKYGGFSQFQEPFSGVGVEVLTPPAVTSIGSTYLIQAOTLTELPNVVR 2775
QY 2348 KALMGKYDOWESPPPLSANAFNPLNASALPAA-MPITAADGRS-----DHTLTSPGGGG 2401
Db 2776 PSDSTHYT--EATPEVSSLA---ELSPQIPSPSPFVYVDNGVSKFPEVPHTSIAQF--- 2825
QY 2402 KAKVSGRPSRKA-----KSPAPGLASGDRPPSVS-----SVHS 2435
Db 2826 ---VSTVTSQKSIESPFEKVEVHANEETIKLGGNVHRTPEPSPMSRDPALDVSDESKHK 2882
QY 2436 EGCNRRRTPLNVRWEDRPSSA-----GSTPPFPYPLNRLQAGVMASPPPLPAGSGP 2490
Db 2883 LLEELSTSPKTPETSQDPFNKAKDHIPGT-----VGLAGIRTTESEPVITADME 2934
QY 2491 LAG-----PHHA 2497
Db 2935 LGGATQOQPSA 2945

RESULT 100
NFH_MOUSE
ID NFH_MOUSE STANDARD; PRT; 1087 AA.
AC F19246; Q61959;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H).
GN NEFH OR NFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eumalia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

